

FIG. 1

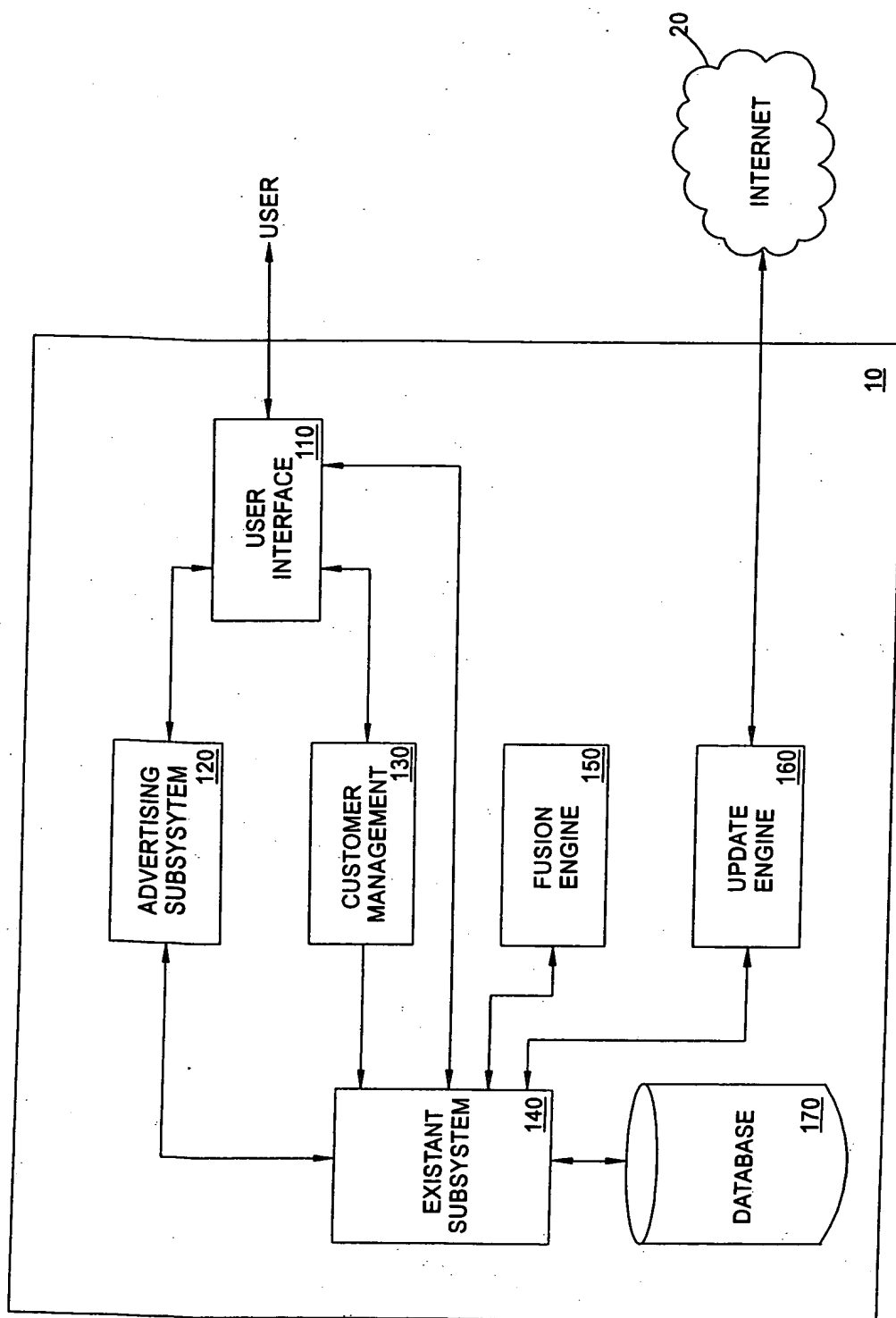


FIG. 2

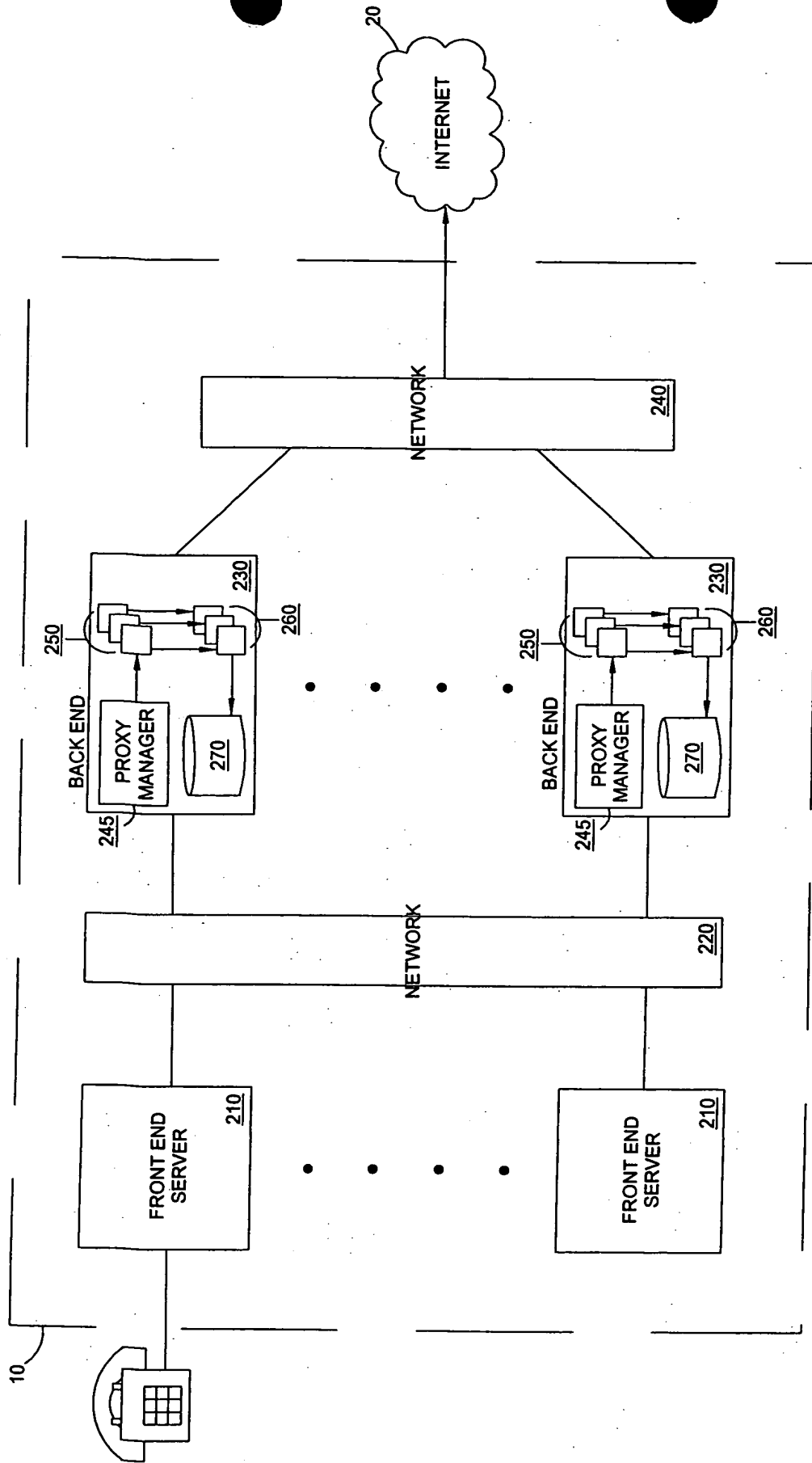


FIG. 3

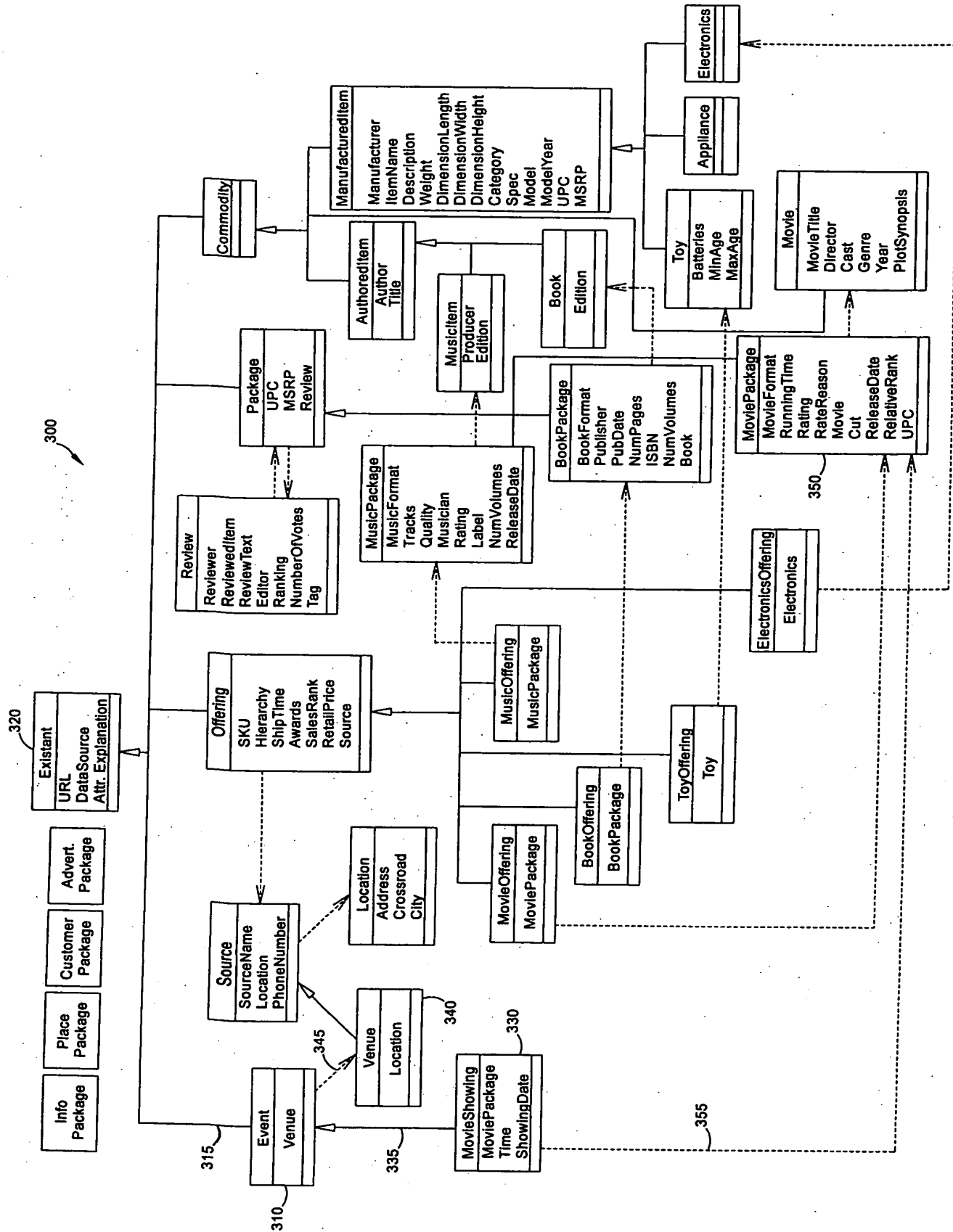


FIG. 4

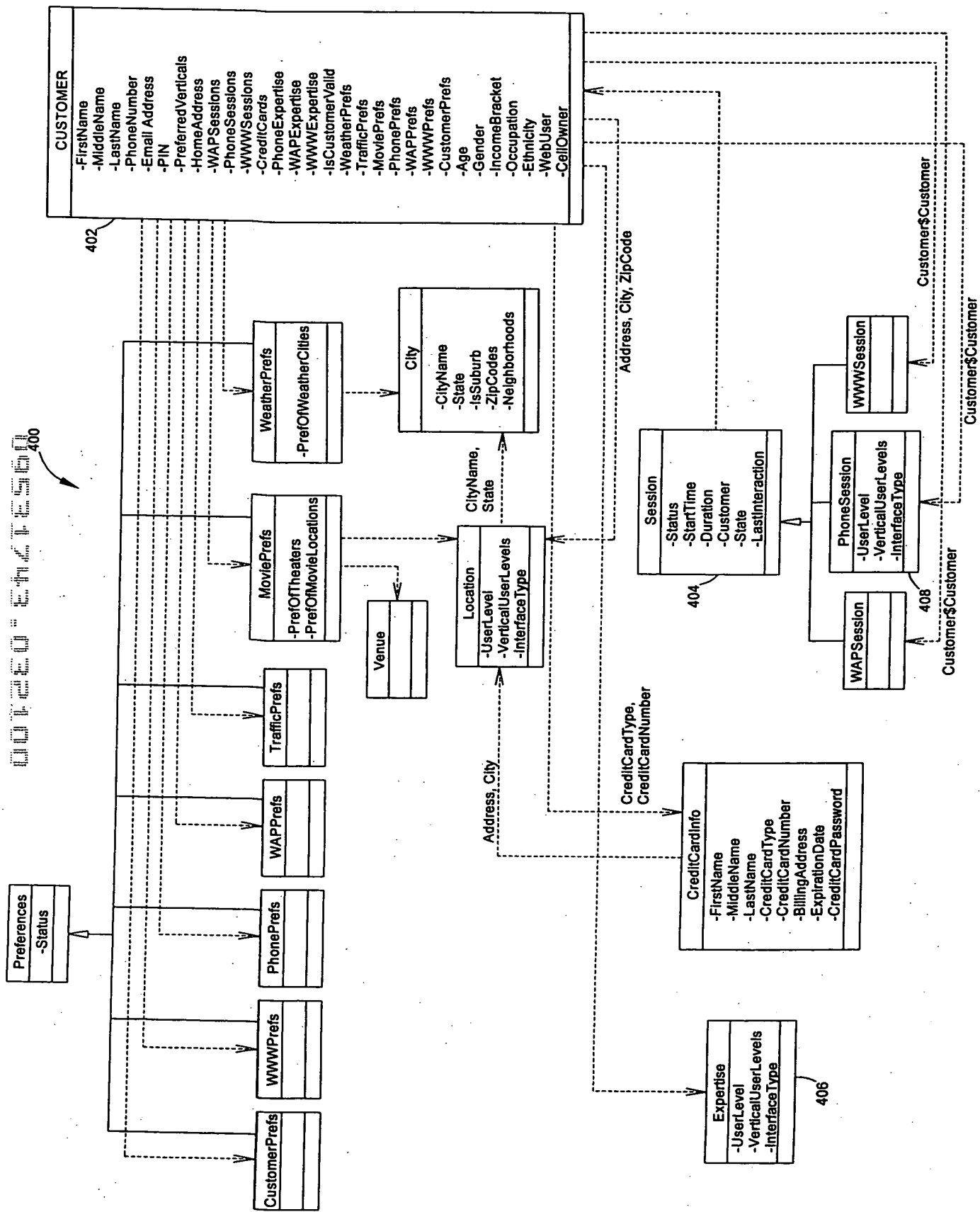


FIG. 5

450

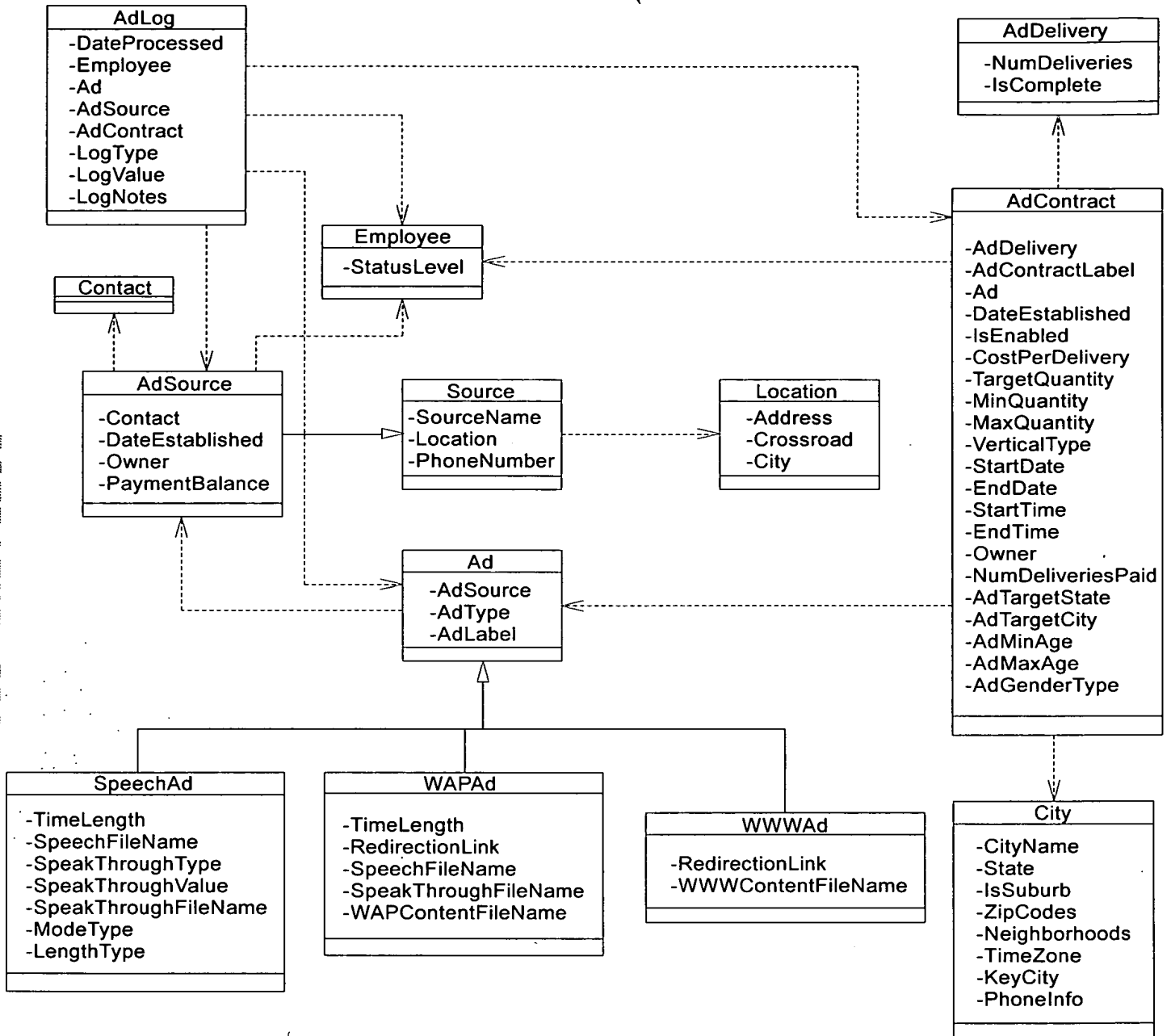


FIG. 6

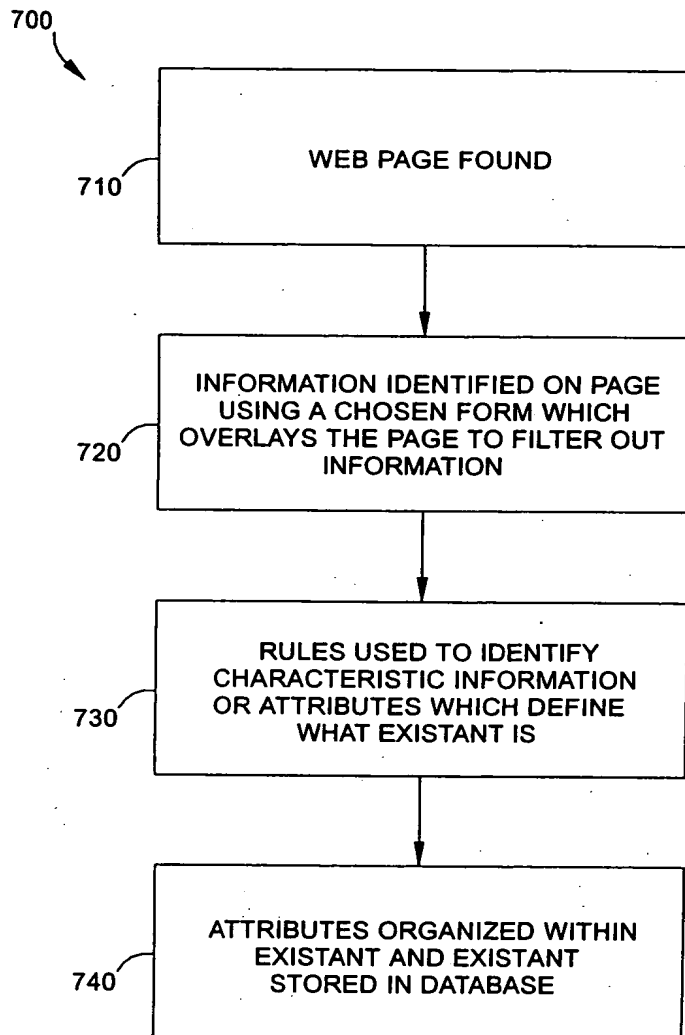


FIG. 7

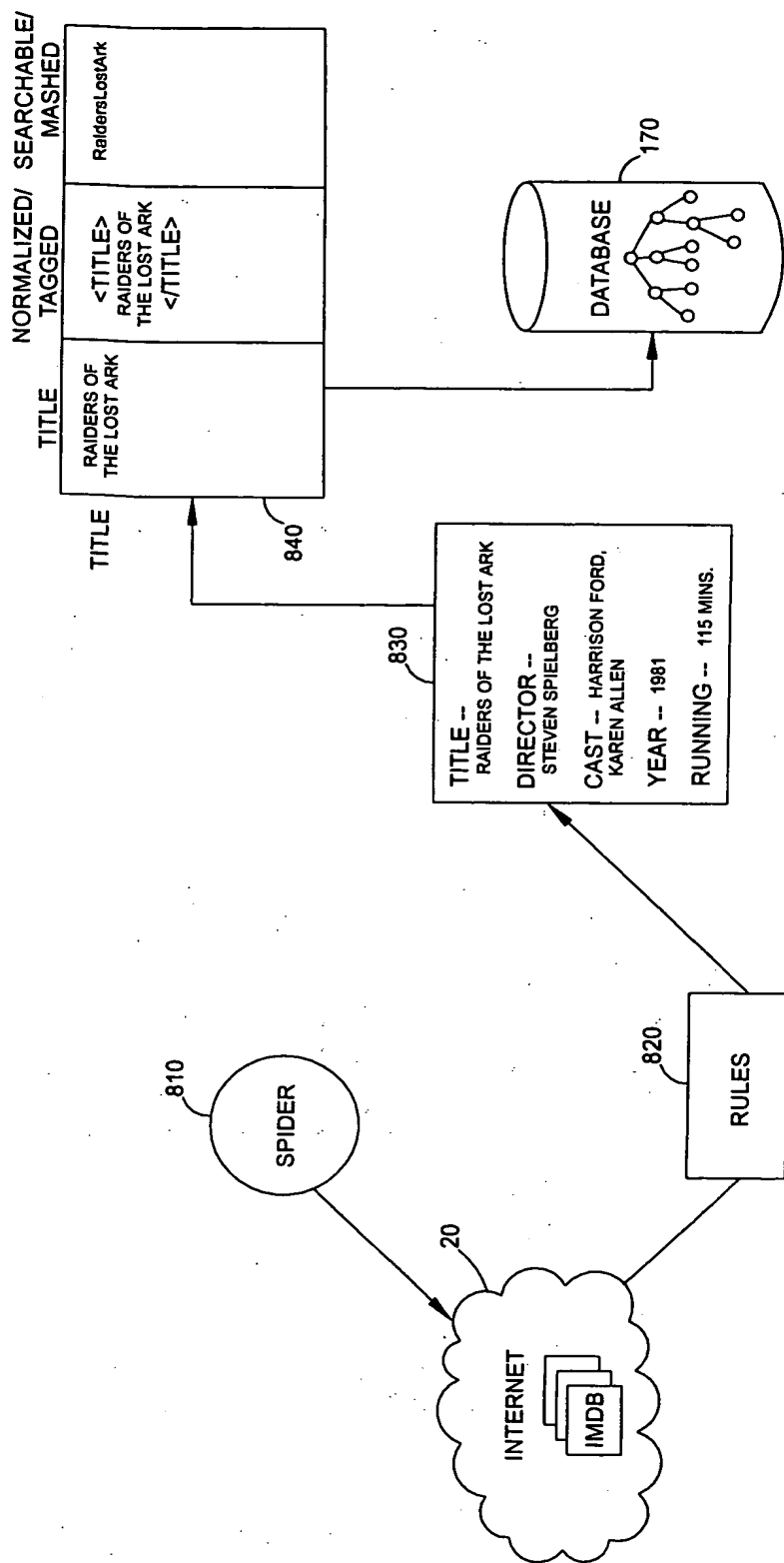


FIG. 8

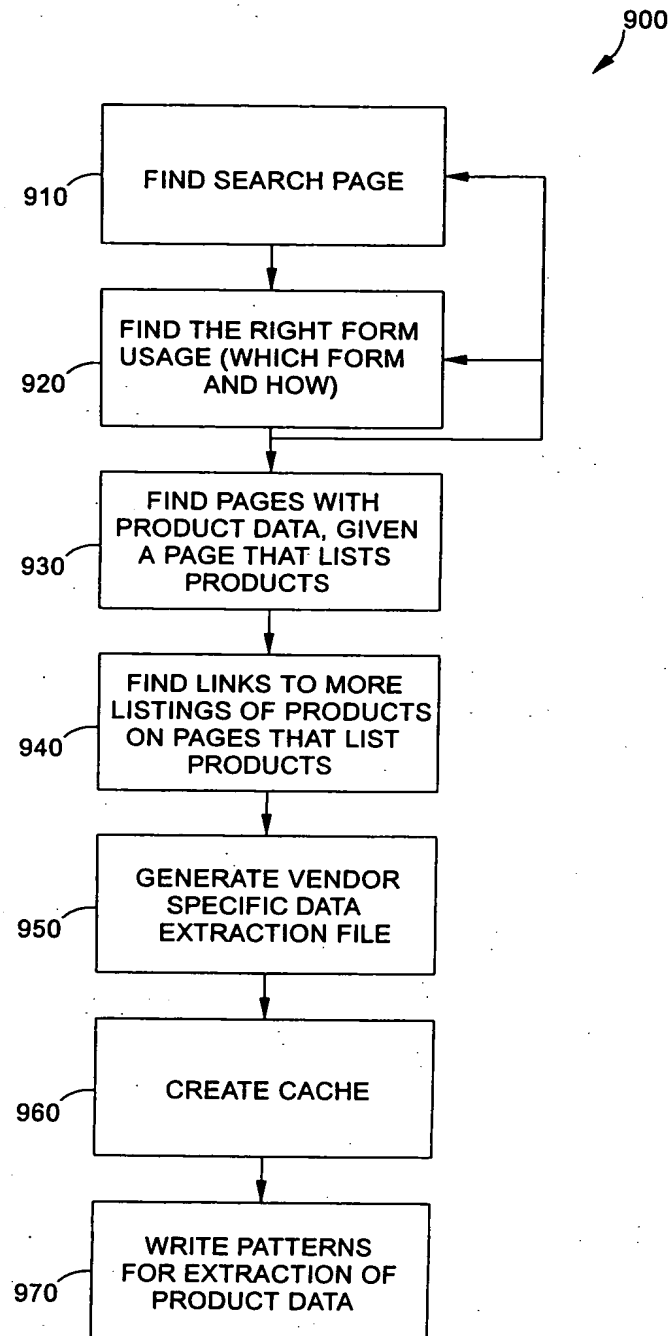


FIG. 9

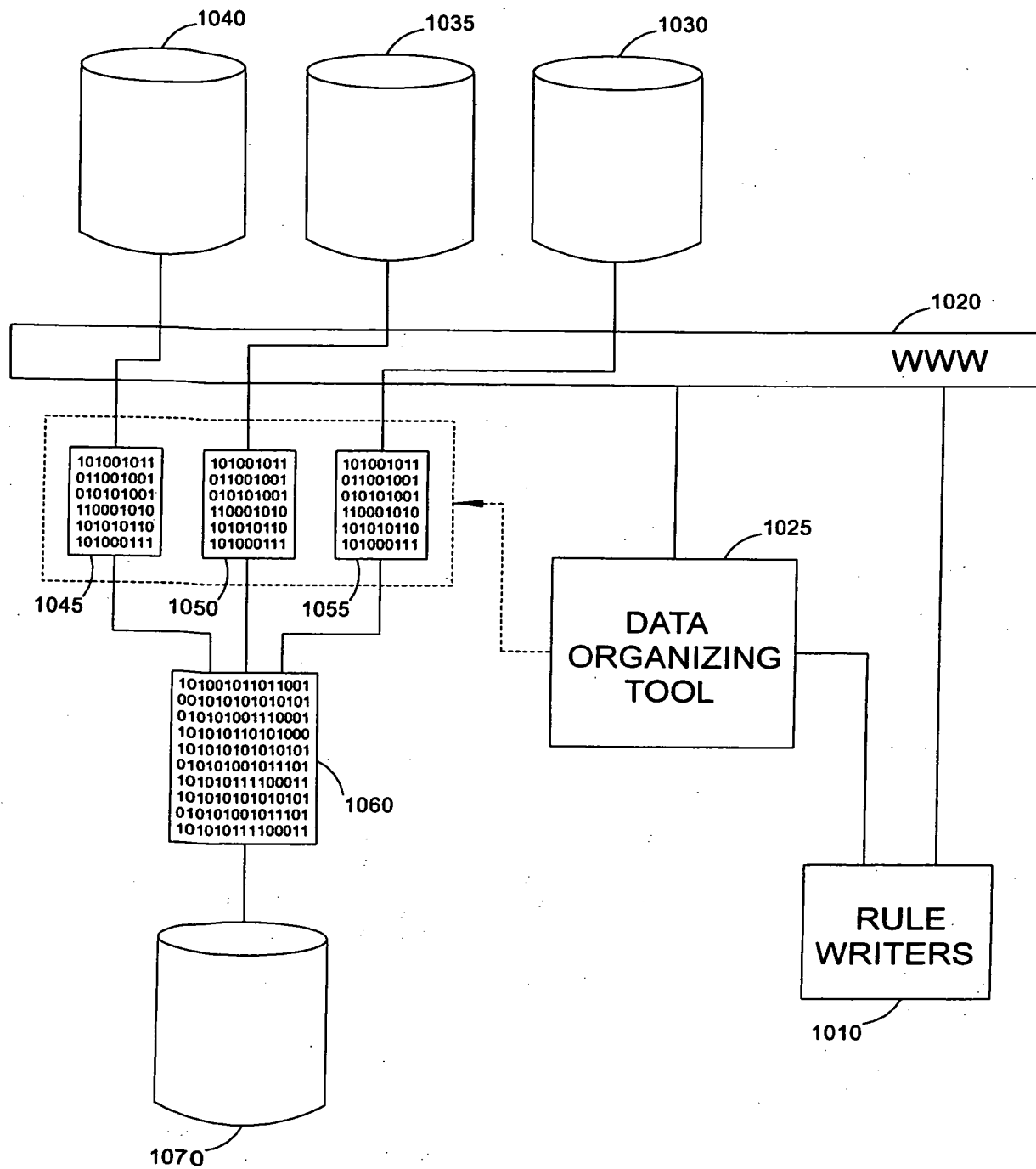


FIG. 10

1200

NewVendor

This routine generates initial versions of all files, needed in the rule writing process. If the file that QUECK wants to generate already exists, a back-up of the original file is saved in /home/karen/.QUECK/Rules.bu or /home/karen/.QUECK/RuleFunctions.bu

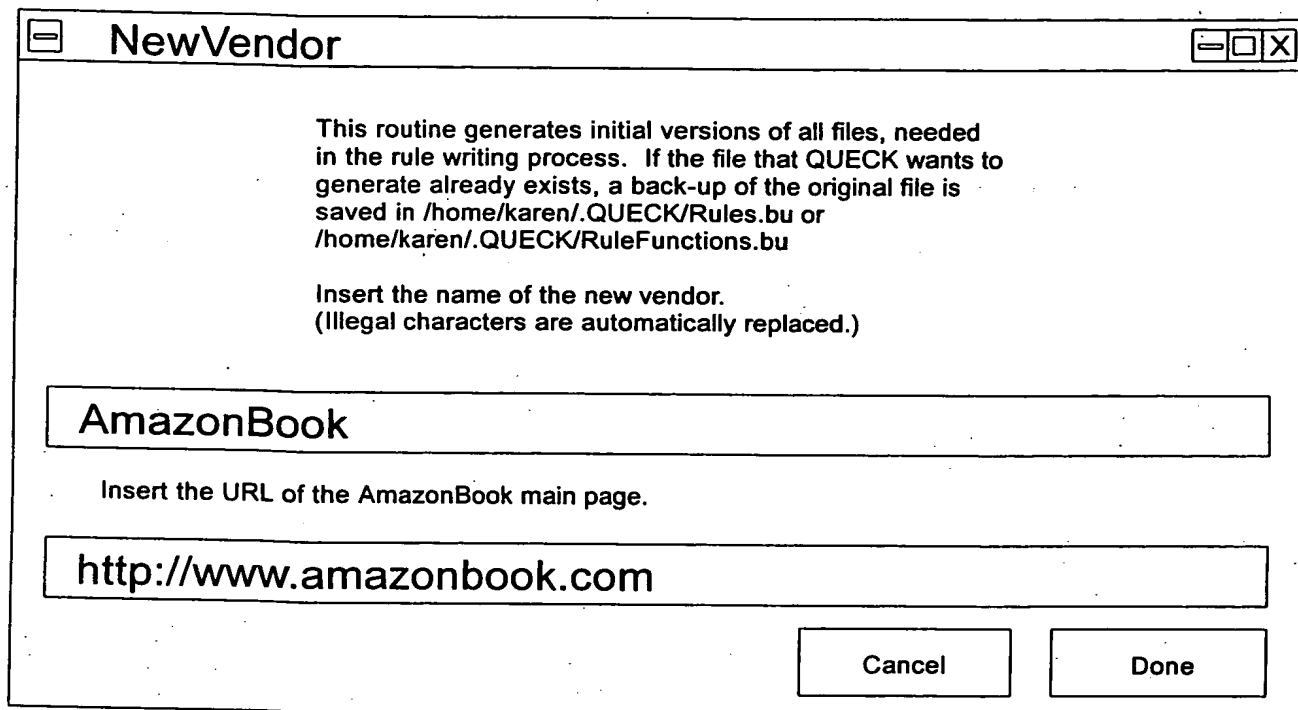
Insert the name of the new vendor.
(Illegal characters are automatically replaced.)

AmazonBook

Cancel Done

FIG. 12

1300



NewVendor

This routine generates initial versions of all files, needed in the rule writing process. If the file that QUECK wants to generate already exists, a back-up of the original file is saved in /home/karen/.QUECK/Rules.bu or /home/karen/.QUECK/RuleFunctions.bu

Insert the name of the new vendor.
(Illegal characters are automatically replaced.)

AmazonBook

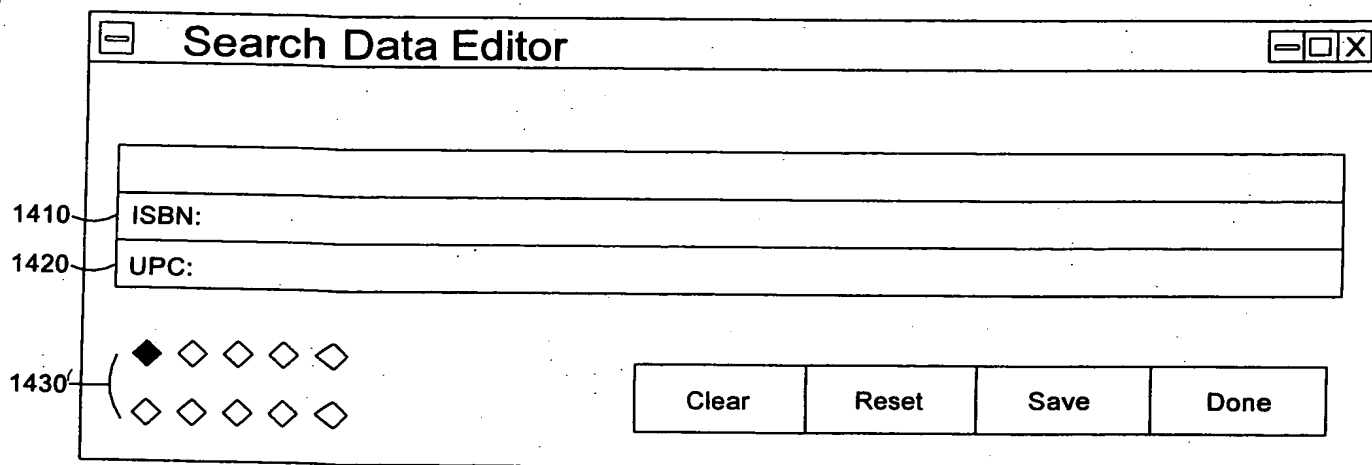
Insert the URL of the AmazonBook main page.

http://www.amazonbook.com

Cancel Done

FIG. 13

1400



Search Data Editor

ISBN:

UPC:

◆ ◇ ◇ ◇ ◇
◇ ◇ ◇ ◇ ◇

Clear Reset Save Done

FIG. 14

1500

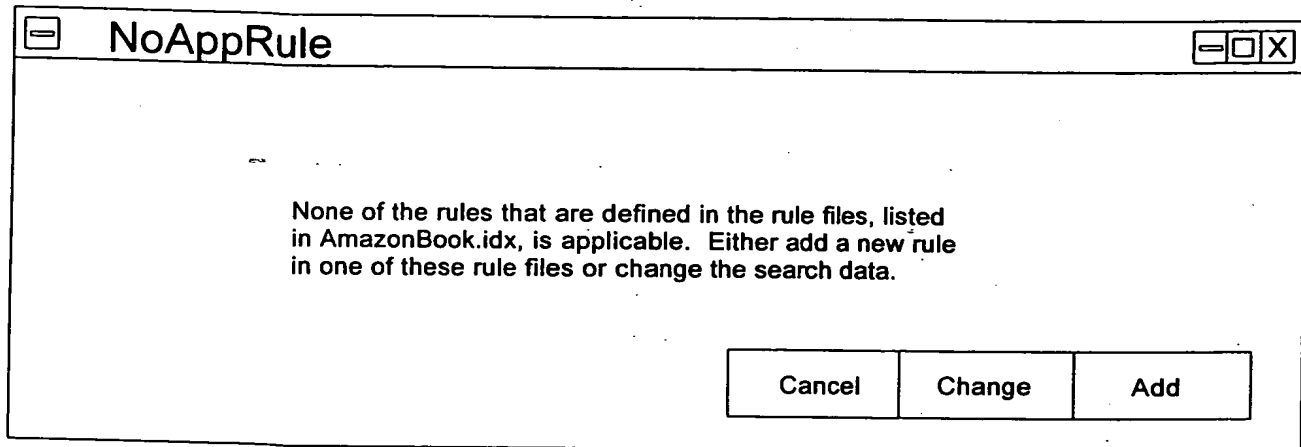


FIG. 15

1600

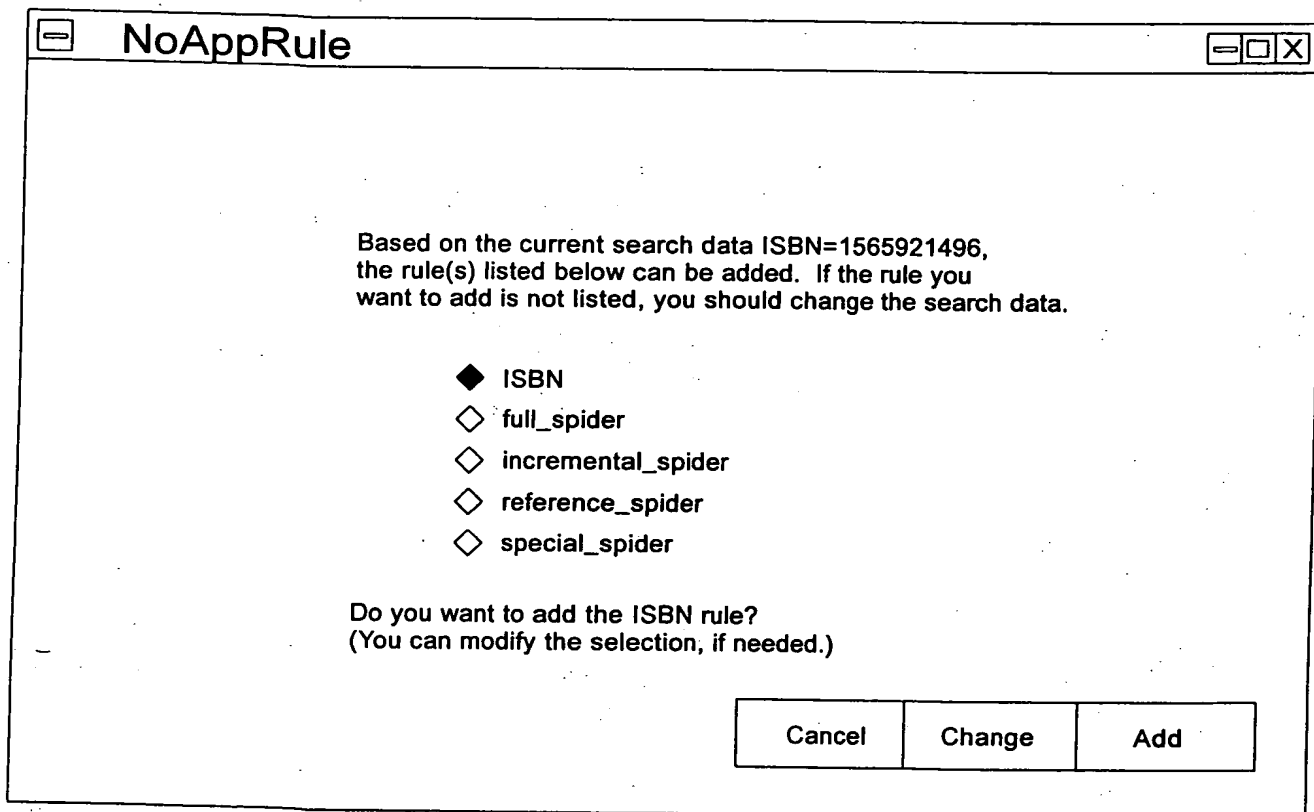


FIG. 16

1700

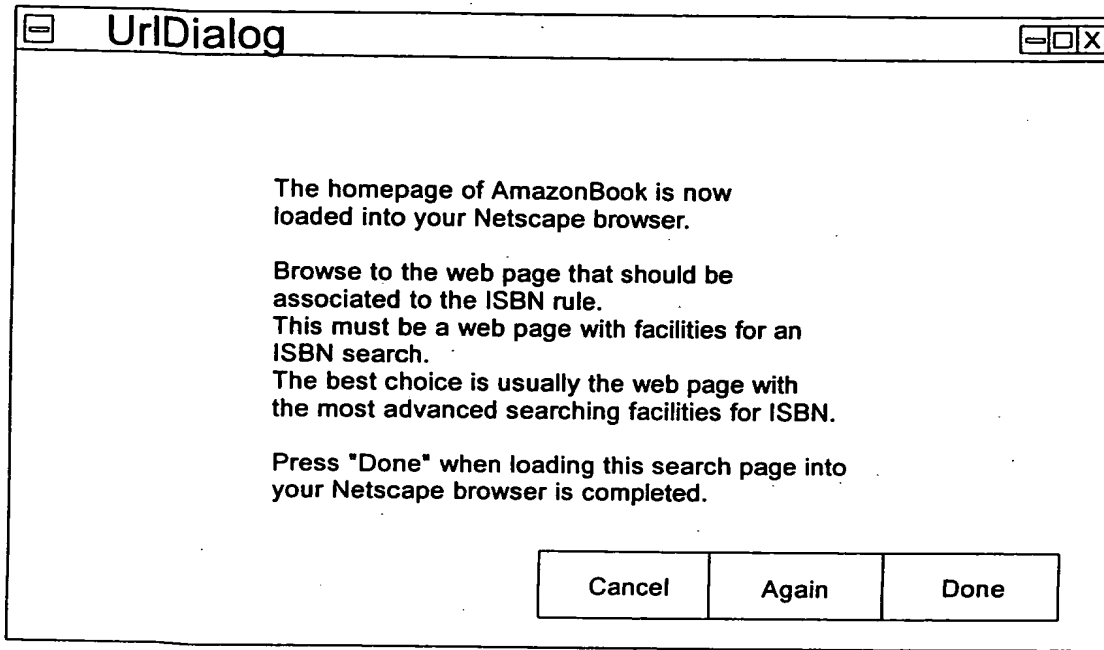


FIG. 17

1800

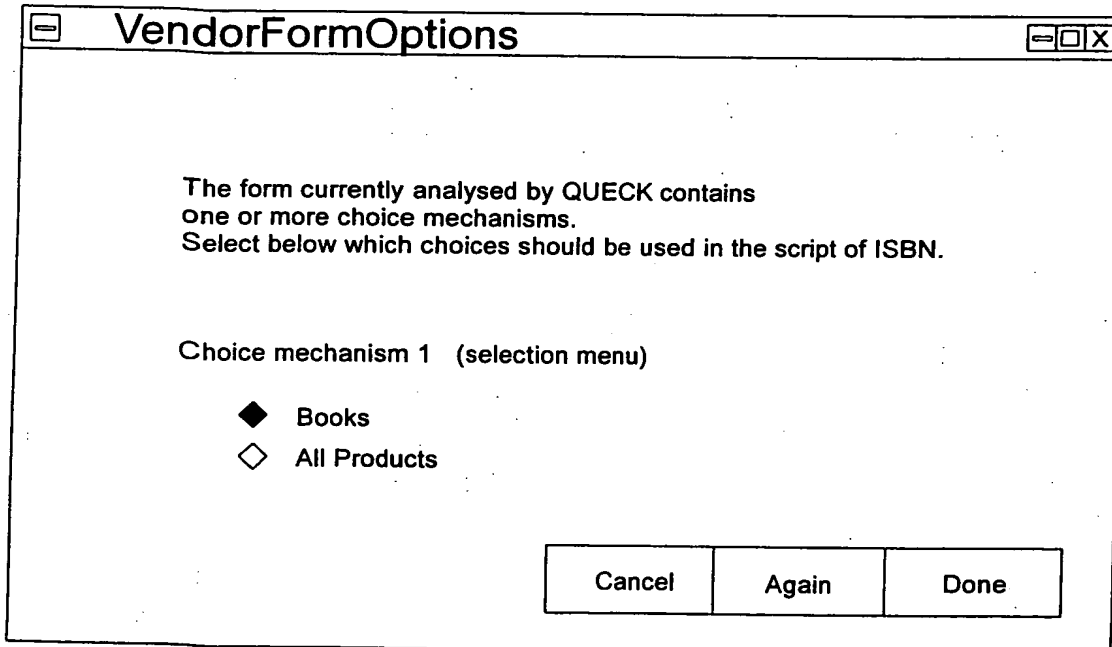
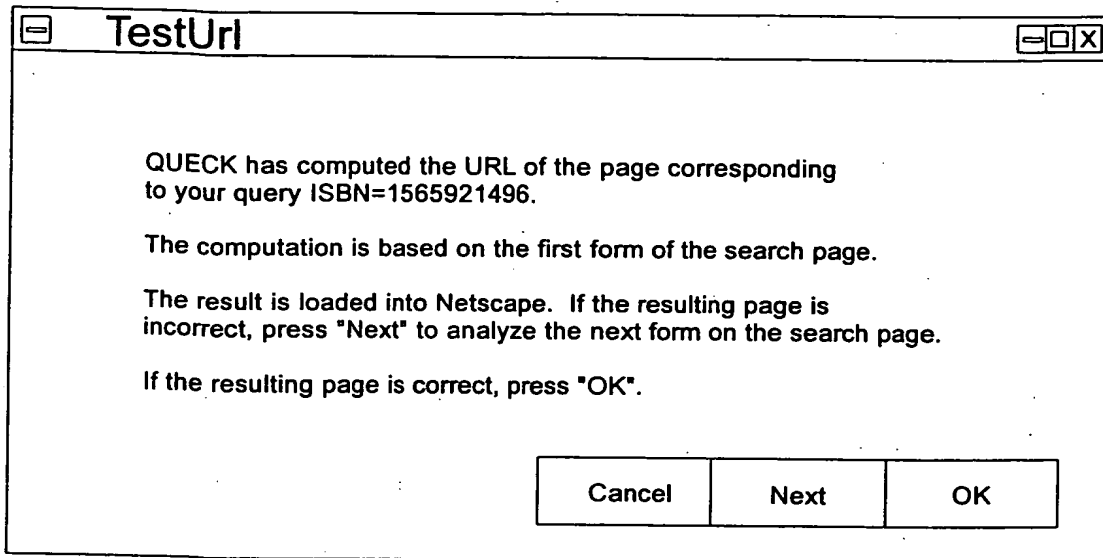


FIG. 18

1900



A dialog box titled "TestUrl" with a standard window control bar (minimize, maximize, close). The text inside reads: "QUECK has computed the URL of the page corresponding to your query ISBN=1565921496. The computation is based on the first form of the search page. The result is loaded into Netscape. If the resulting page is incorrect, press 'Next' to analyze the next form on the search page. If the resulting page is correct, press 'OK'." At the bottom right, there are three buttons: "Cancel", "Next", and "OK".

TestUrl

QUECK has computed the URL of the page corresponding to your query ISBN=1565921496.

The computation is based on the first form of the search page.

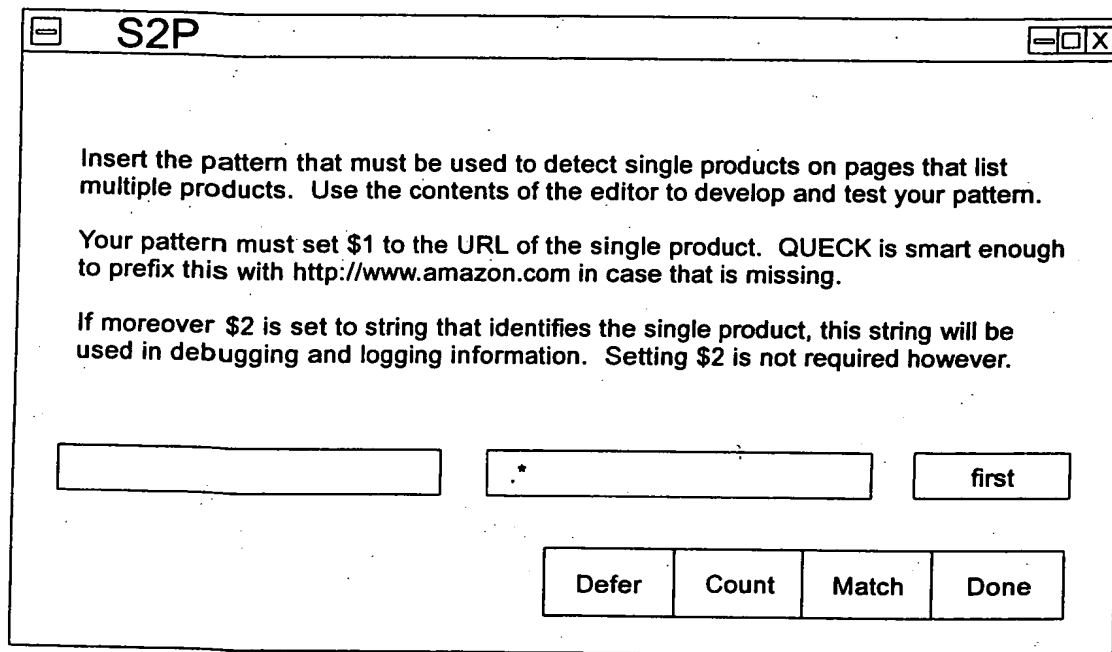
The result is loaded into Netscape. If the resulting page is incorrect, press "Next" to analyze the next form on the search page.

If the resulting page is correct, press "OK".

Cancel Next OK

FIG. 19

2000



A dialog box titled "S2P" with a standard window control bar (minimize, maximize, close). The text inside reads: "Insert the pattern that must be used to detect single products on pages that list multiple products. Use the contents of the editor to develop and test your pattern. Your pattern must set \$1 to the URL of the single product. QUECK is smart enough to prefix this with http://www.amazon.com in case that is missing. If moreover \$2 is set to string that identifies the single product, this string will be used in debugging and logging information. Setting \$2 is not required however." Below the text are two input fields: the first is empty, and the second contains an asterisk (*). To the right of the second input field is a button labeled "first". At the bottom right, there are four buttons: "Defer", "Count", "Match", and "Done".

S2P

Insert the pattern that must be used to detect single products on pages that list multiple products. Use the contents of the editor to develop and test your pattern.

Your pattern must set \$1 to the URL of the single product. QUECK is smart enough to prefix this with http://www.amazon.com in case that is missing.

If moreover \$2 is set to string that identifies the single product, this string will be used in debugging and logging information. Setting \$2 is not required however.

first

Defer Count Match Done

FIG. 20

2100

NSP

Insert the pattern that must be used to detect links on multiple products pages to even more multiple product pages. Use the "Match" button to test your pattern.

Your pattern must set \$1 to the URL of the new multiple product page. QUECK is smart enough to prefix this with http://www.amazon.com in case that is missing.

If your query does not generate enough product hits to have more than one multiple product page, you can choose "Defer" and defer the configuration until you run a query that actually does generate enough product hits to have more than one multiple product page.

Defer

Match

Build

FIG. 21

```

graph TD
    2210(( )) --- 2220(( ))
    2210 --- 2211(( ))
    2210 --- 2212(( ))
    2220 --- 2230(( ))
    2220 --- 2221(( ))
    2230 --- 2240(( ))
    2230 --- 2231(( ))
    2230 --- 2232(( ))
    2230 --- 2233(( ))
    2221 --- 2222(( ))
    2221 --- 2223(( ))
    2212 --- 2213(( ))
    2212 --- 2214(( ))
    2212 --- 2215(( ))
    2212 --- 2216(( ))
    2213 --- 2217(( ))
    2213 --- 2218(( ))
    2213 --- 2219(( ))
    2213 --- 2220(( ))
  
```

FIG. 22

2300

SpiderSubr

Insert here the URL of the page, currently loaded into Netscape. This is the page associated to the full_spider rule.

Next, set "SpiderDepth" to the maximum number of links that has to be followed from the top of the hierarchy to the actual product pages. Note that in some cases this number depends on the branch you follow. Setting "SpiderDepth" too low creates a spider that misses products that are nested too deep in the hierarchy. Setting "SpiderDepth" too high leads to a decrease in performance.

SpiderDepth

1

UpperBound

0

Done

FIG. 23

2400

SpiderSubr

http://www.amazon.com/exec/obidos/subst/home/home.html/002-5797861-2625002

The spider you specified is a level - 1 spider.
This means that your spider has the following form:

level - 0: The top page (accessed via the URL above)

level - 1: The single product pages to be spidered

Insert below the pattern used to detect level - 1 pages on the top page.

Your pattern must set \$1 to the URLs of the child pages. QUECK is smart enough to prefix this URL with http://www.amazon.com in case it is missing. If your pattern also sets \$2, that value will be used in the hierarchy attributes.

first

1st Level

.

Cancel

Count

Match

Build

FIG. 24

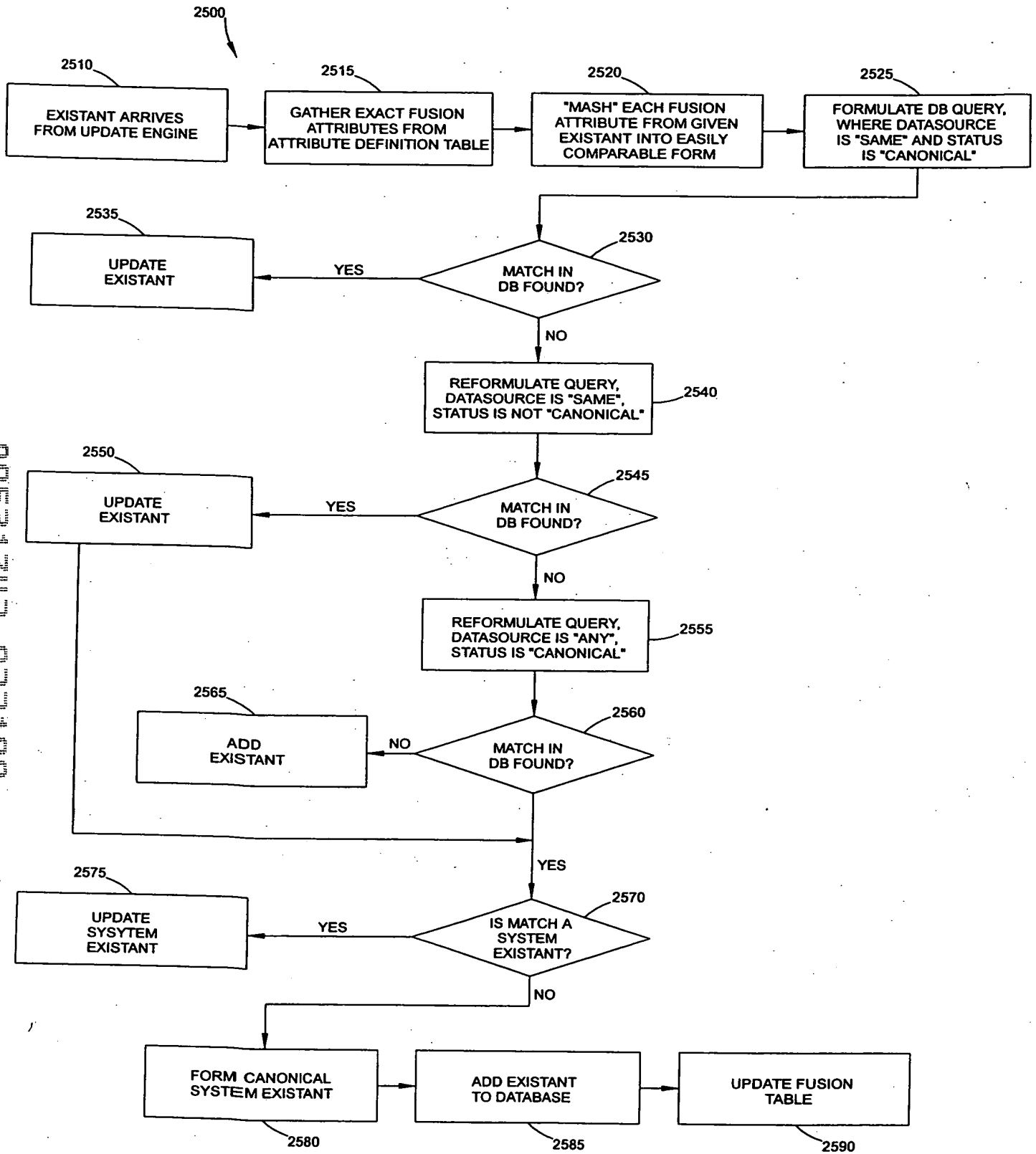


FIG. 25

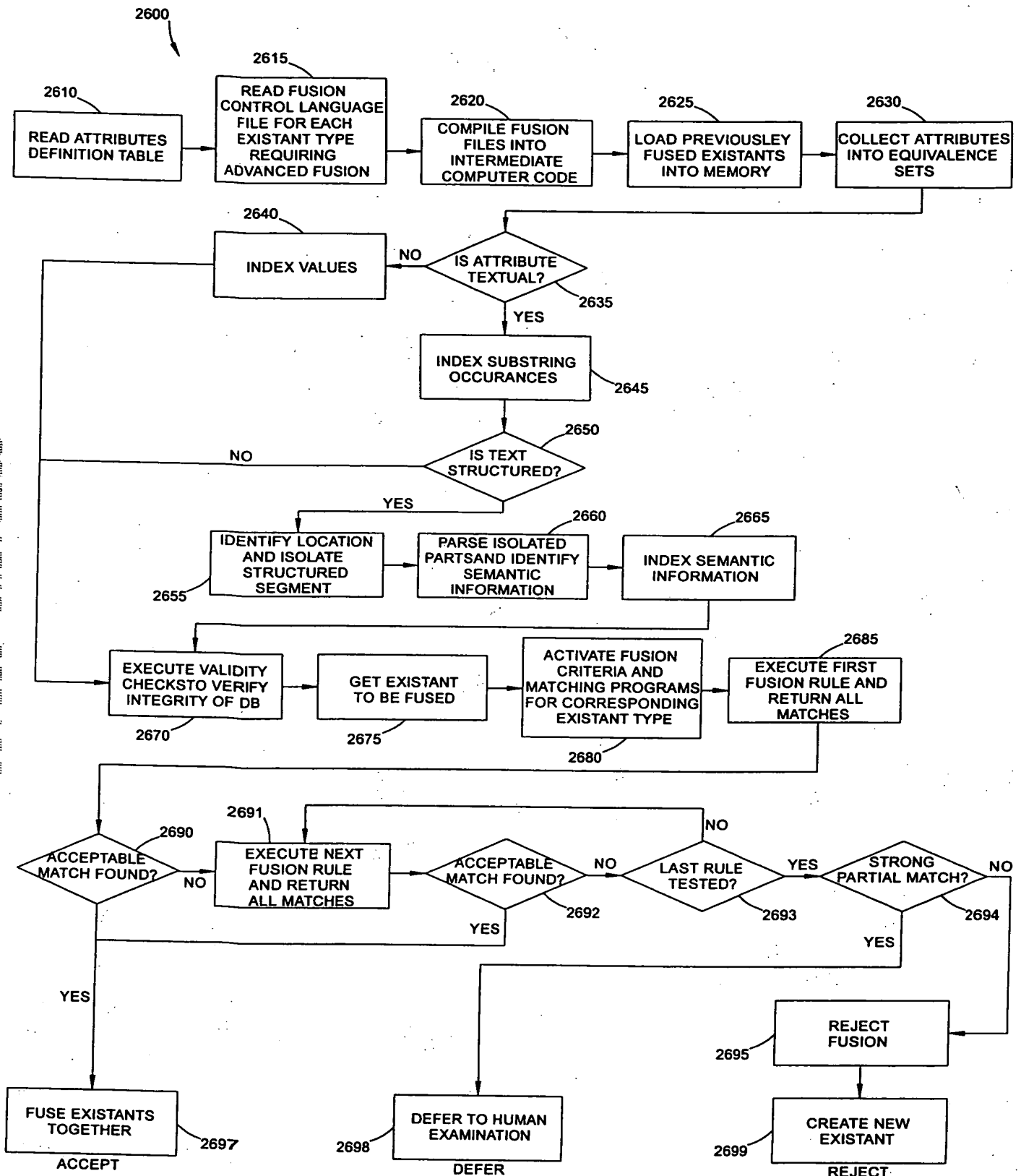


FIG. 26

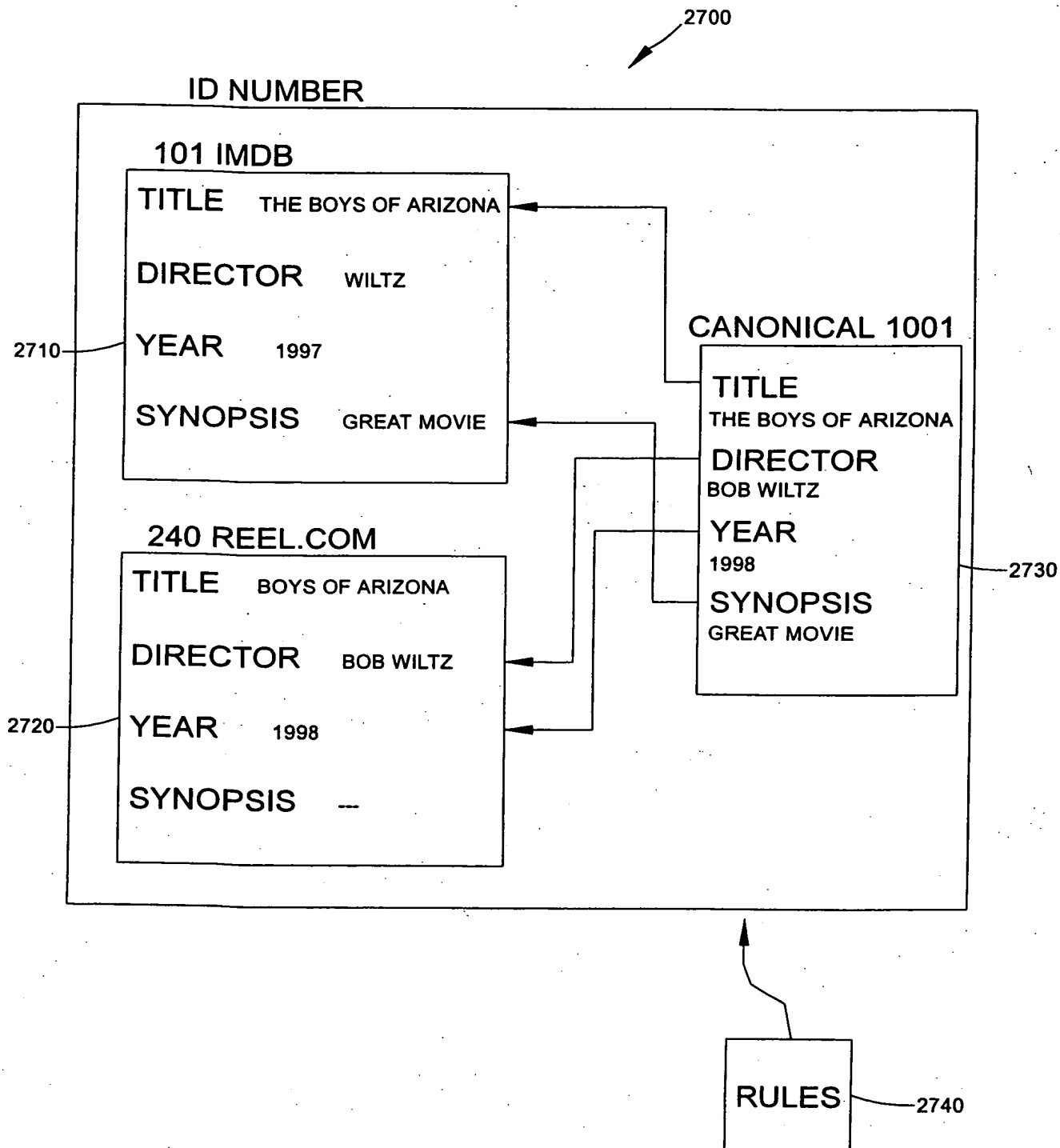


FIG. 27

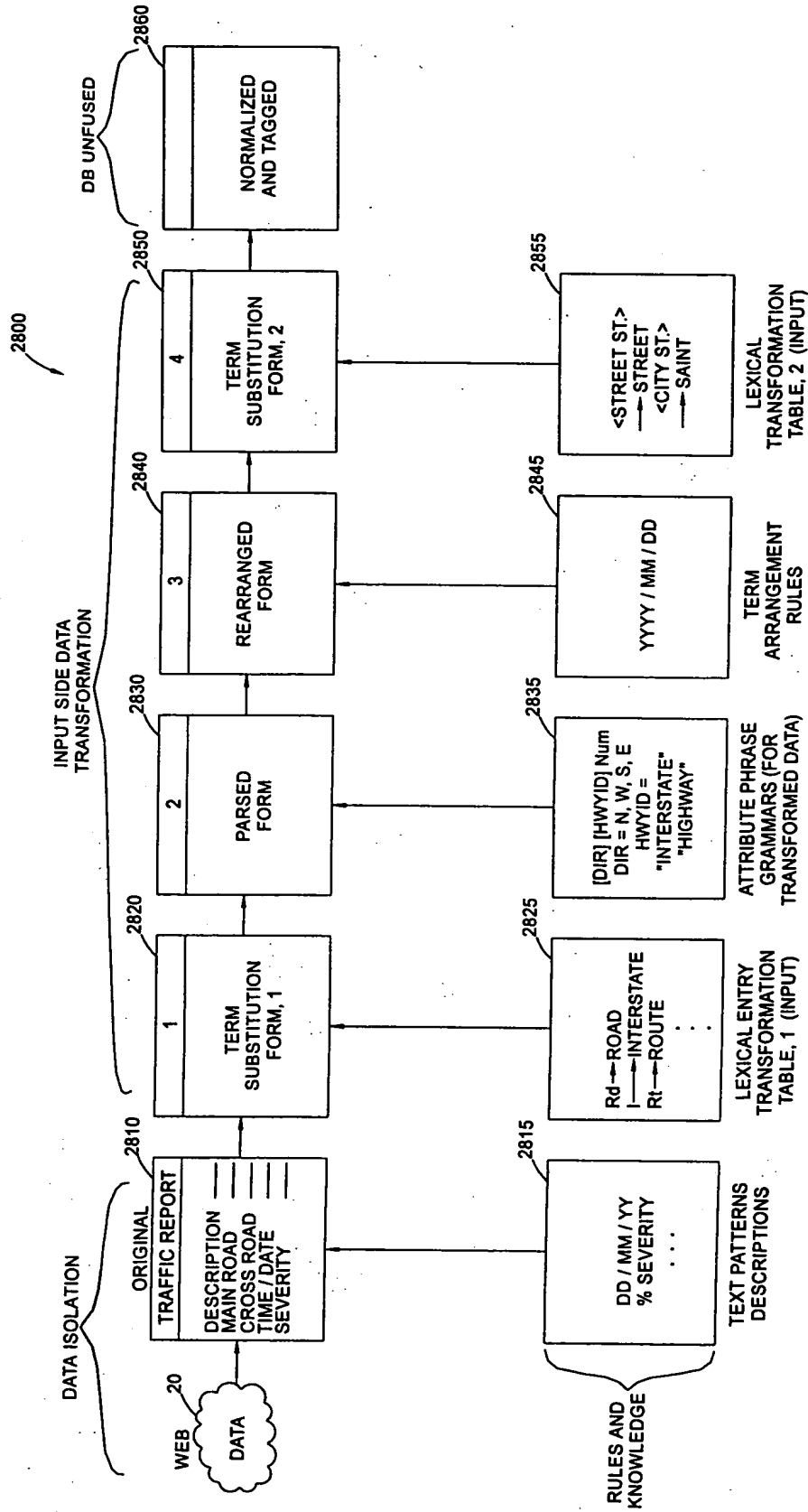


FIG. 28

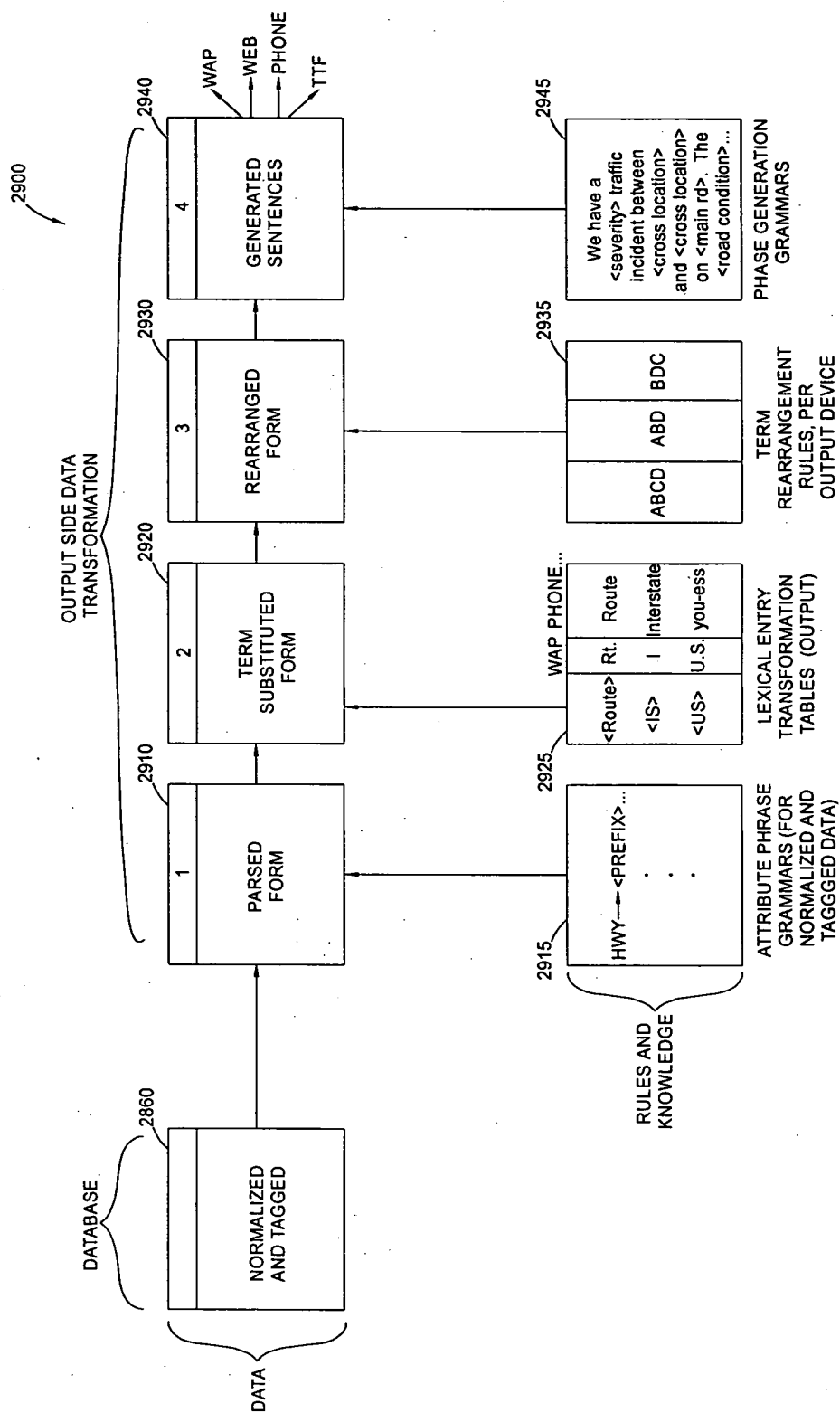


FIG. 29

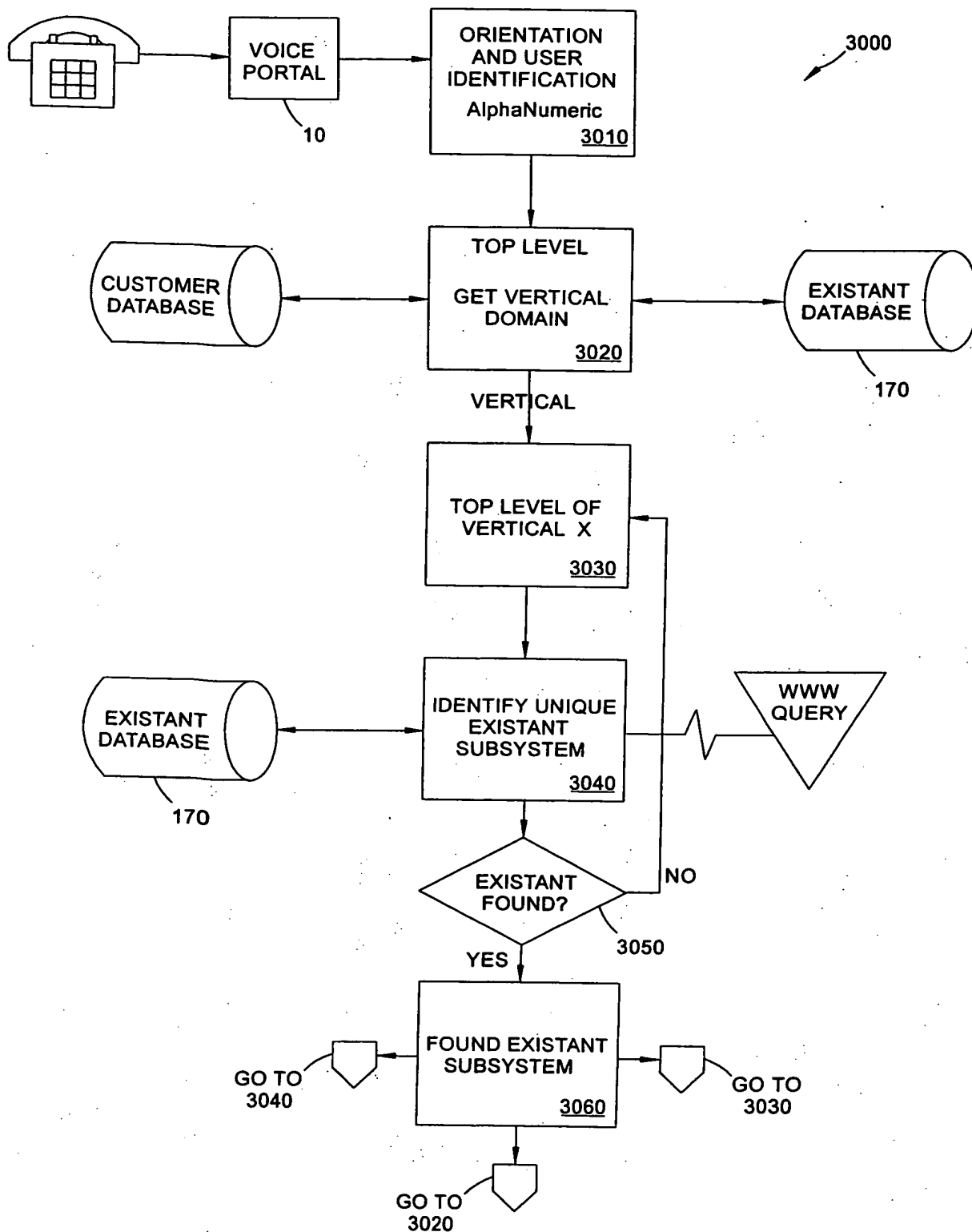


FIG. 30


```

graph TD
    3040[3040] --> 3110[3110]
    3110 --> 3115[3115]
    3115 --> 3030[3030]
    3110 --> 3120[3120]
    3120 --> 3130[3130]
    3130 --> 3140{3140}
    3140 -- YES --> 3150[3150]
    3140 -- NO --> 3110
    3150 --> 3200[3200]

```

FIG. 31

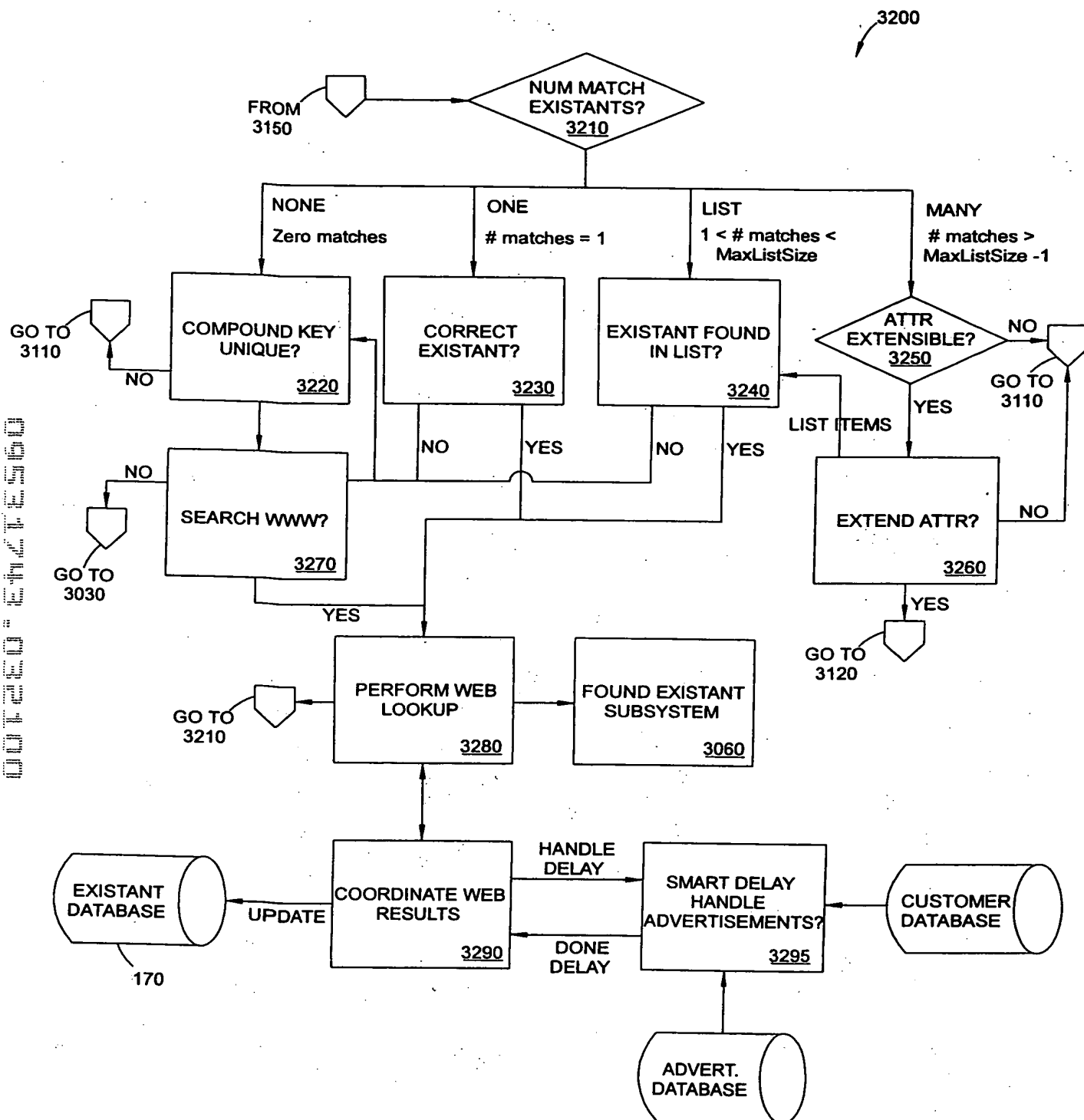


FIG. 32

001220-2421650

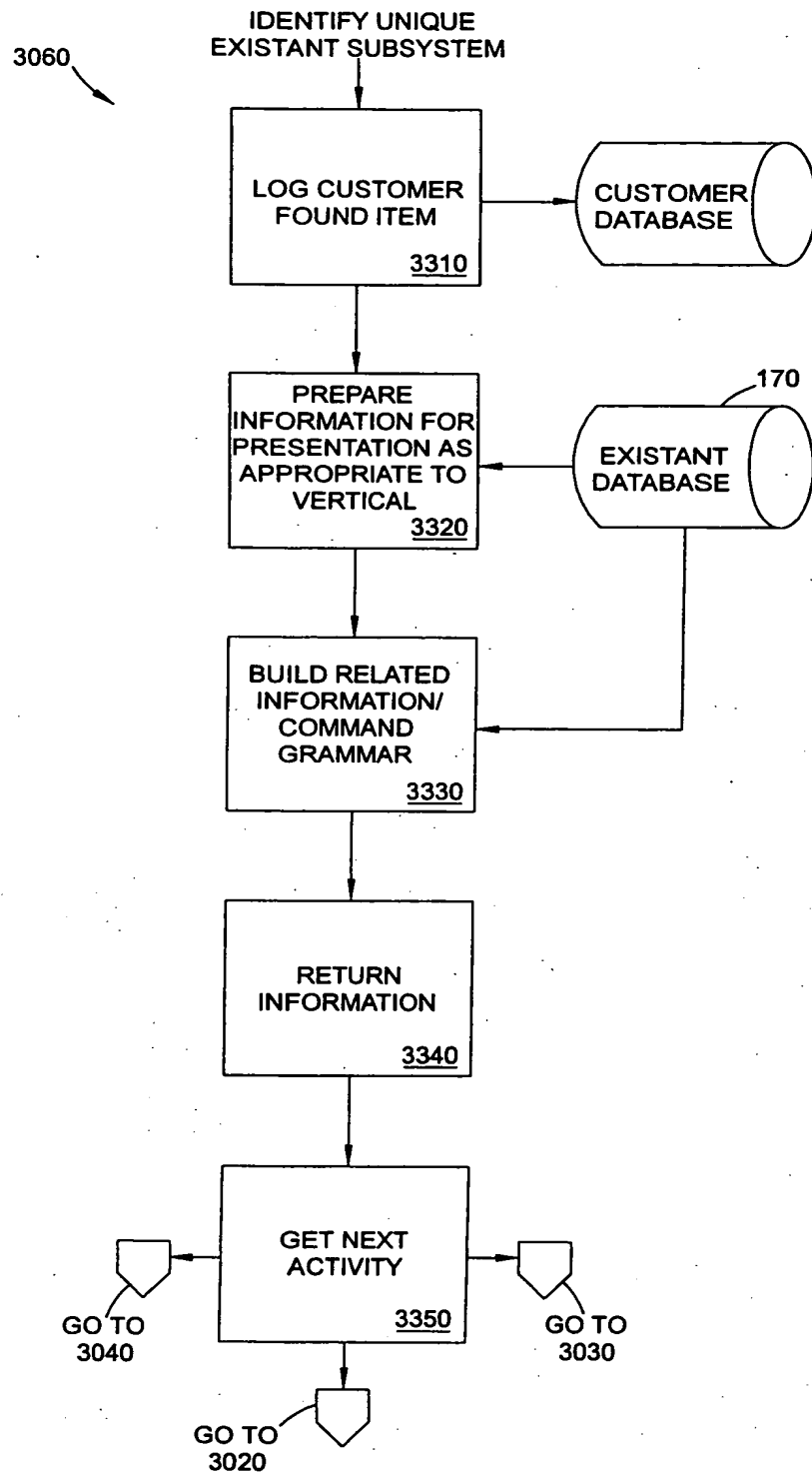


FIG. 33

3400

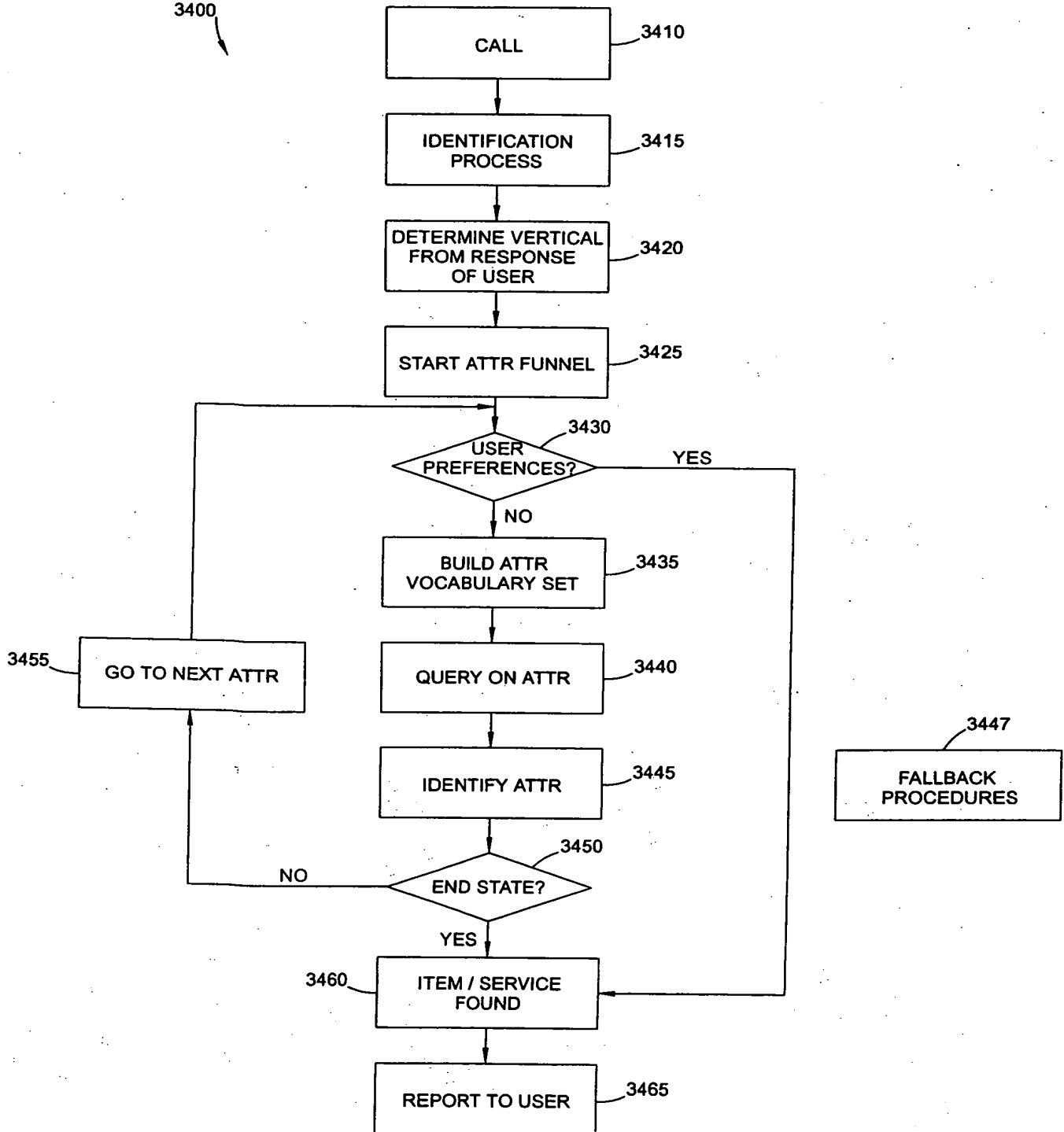


FIG. 34

09591743-032100

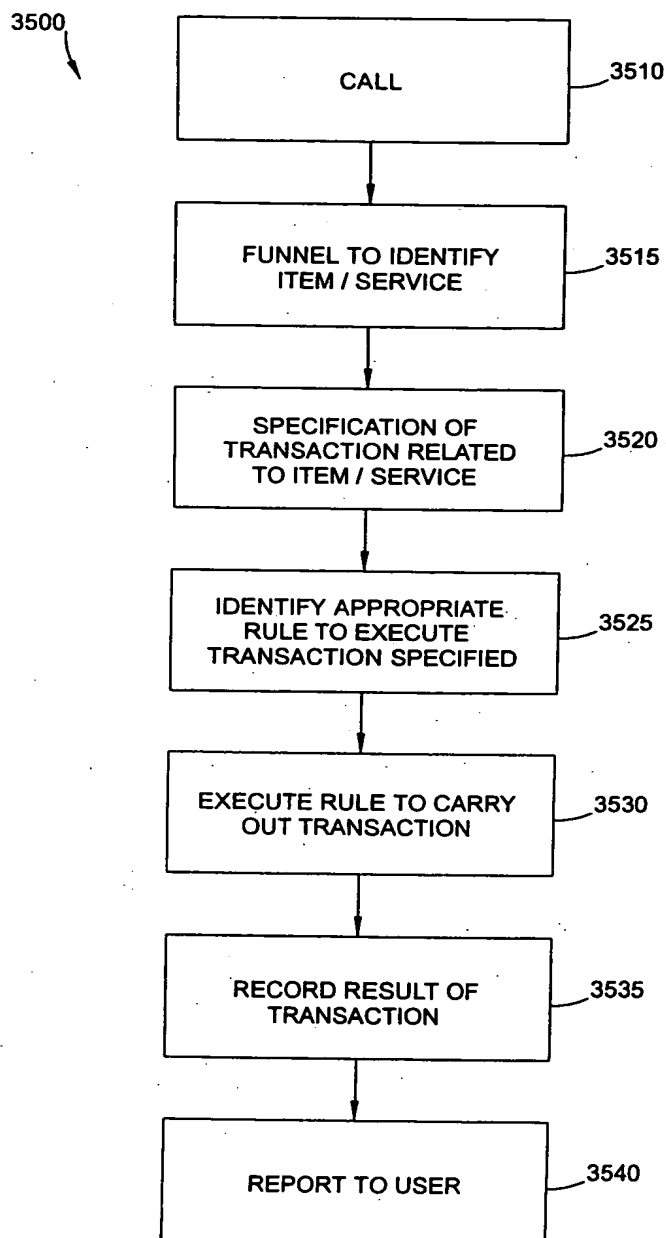


FIG. 35

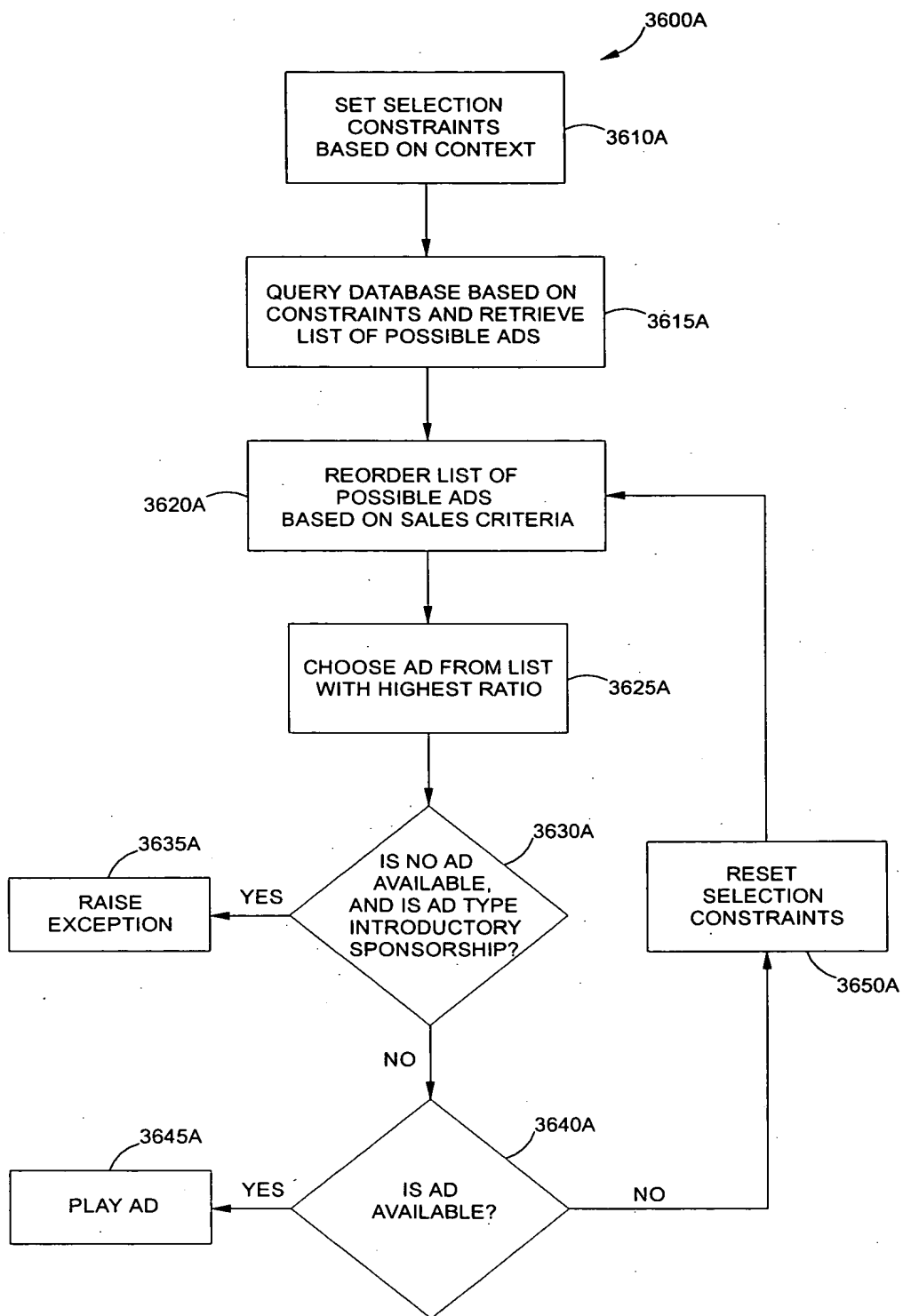


FIG. 36A

001260-ETH2E5B0

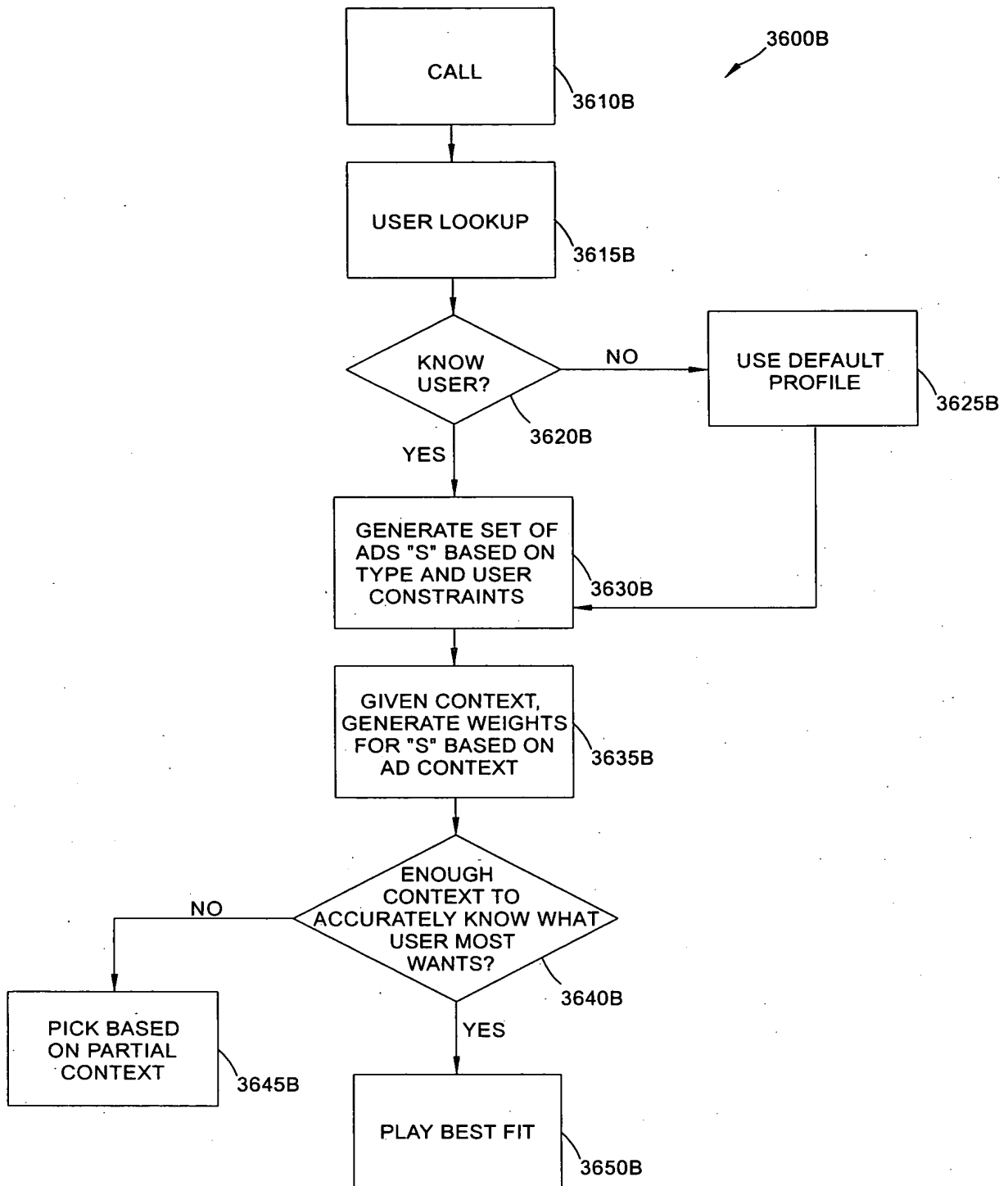


FIG. 36B

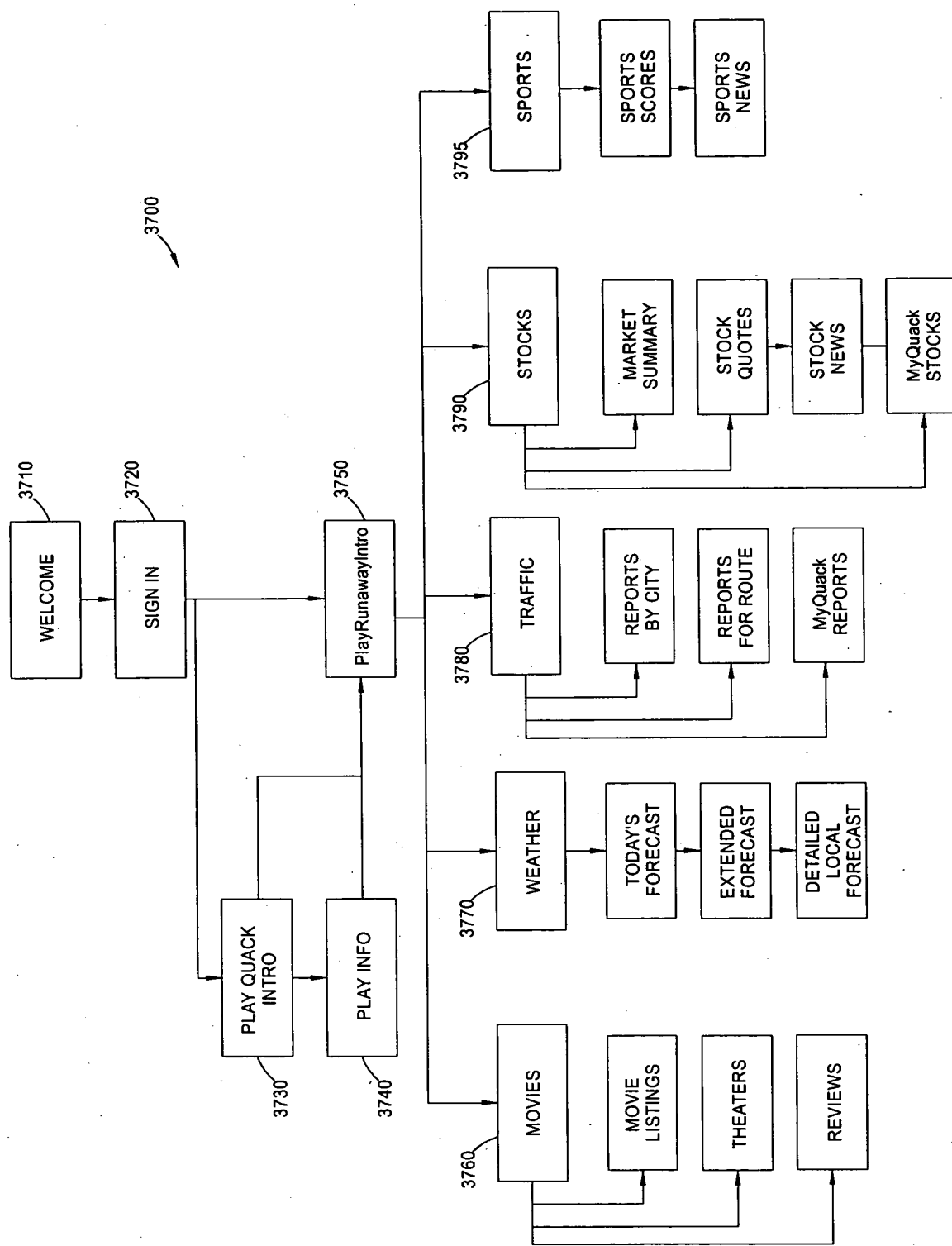


FIG. 37

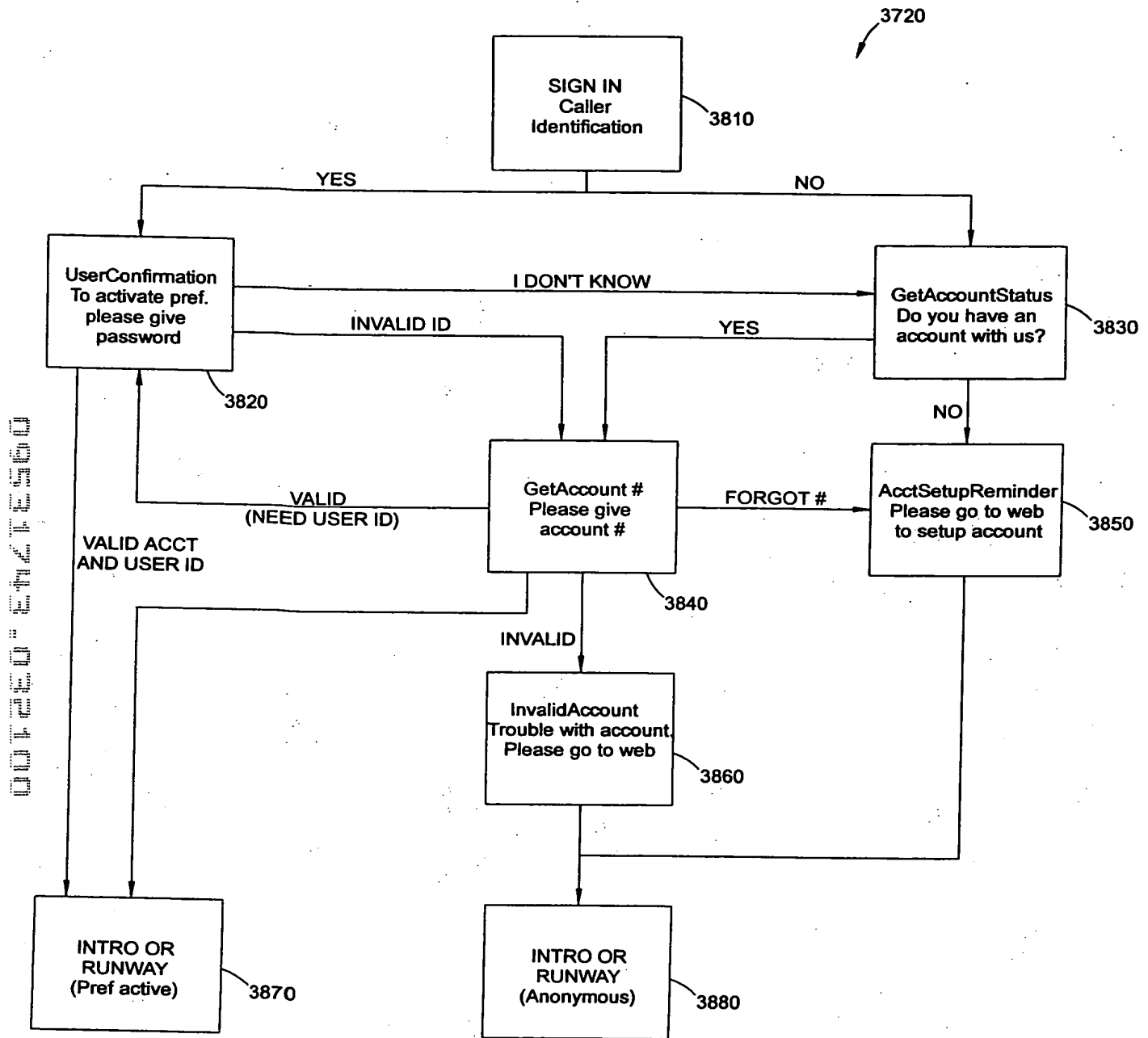


FIG. 38

001220" 6421650

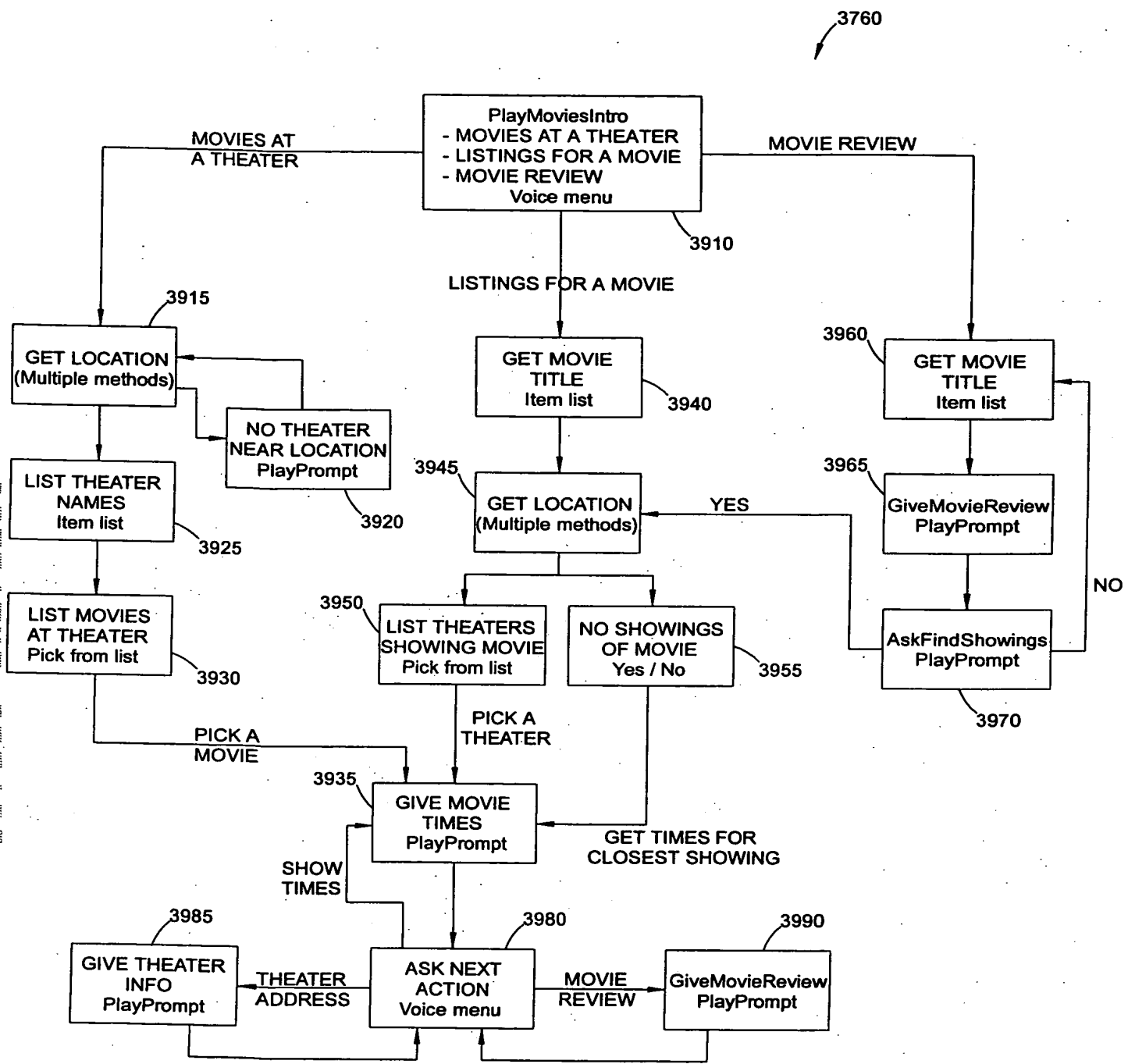


FIG. 39

00000-CH47E560

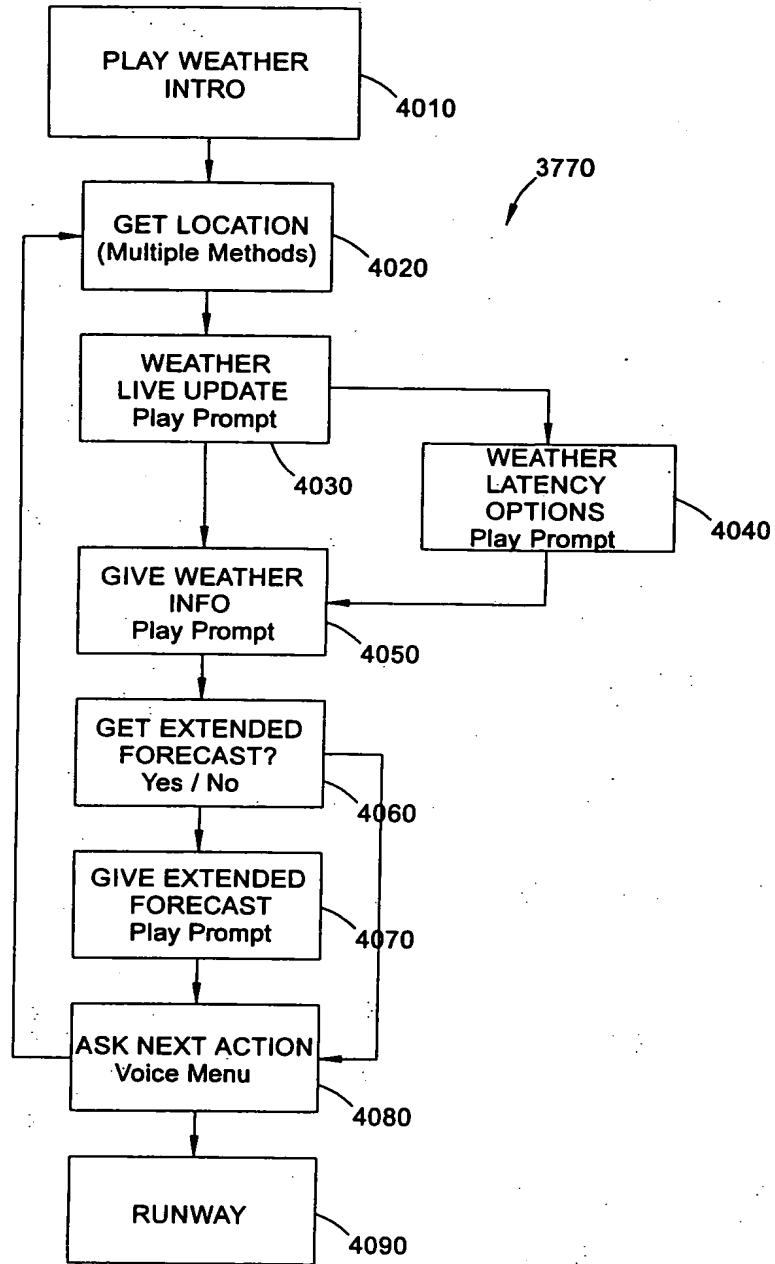


FIG. 40

4180'

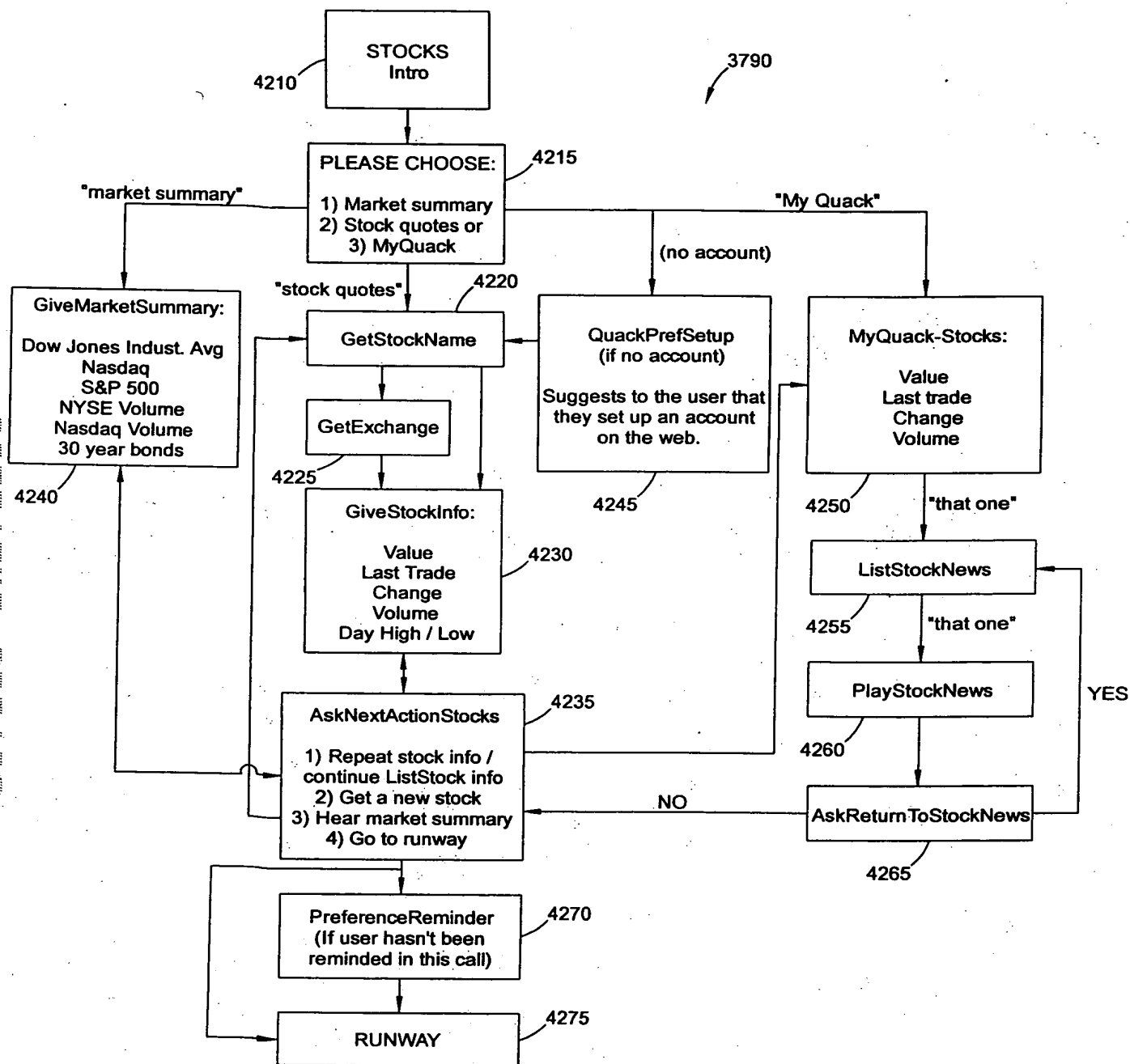


FIG. 42

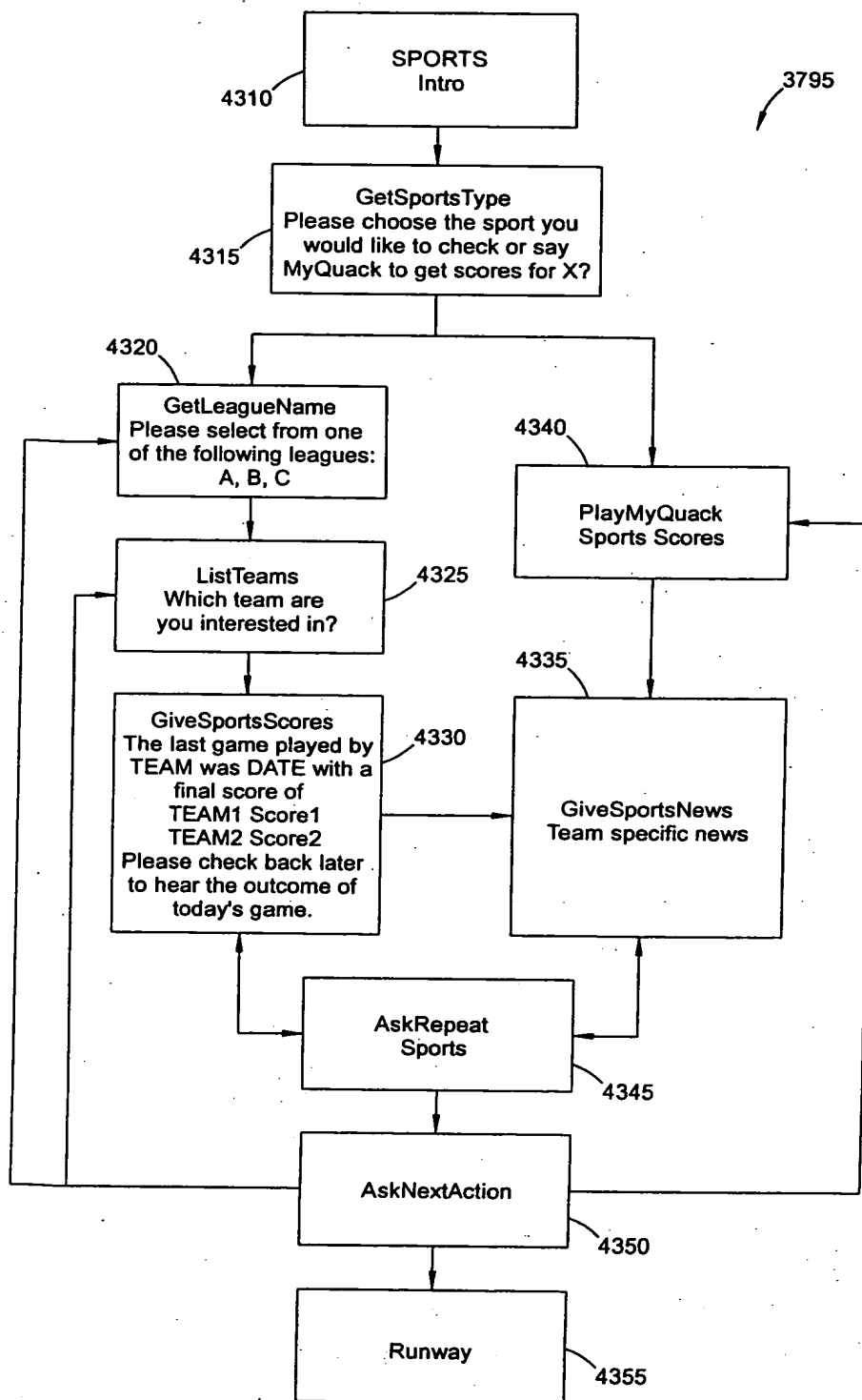


FIG. 43